THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 41, PART 4 | *June* 15, 1929 | NO. 229

THE NEW ACOUSTICS

Presidential address delivered by

W. H. ECCLES, D.Sc., F.R.S.

March 22, 1929

It is just over fifty years since Lord Rayleigh published his *Theory of Sound*. Although it was mainly a mathematical treatment of the subject it comprised much experimental matter and was undoubtedly the most comprehensive work on acoustics that had till then been published. It constituted a treatise of the whole physical theory and in it was arranged in logical order a great mass of material collected from all sources. Many old lacunae in the subject were filled by the author's own original work, experimental and mathematical; and as a consequence we may take it as representing the whole of the physical acoustics of fifty years ago.

The second edition appeared about eighteen years later, and was more than half as large again. Much of the new matter was original, especially that relating to the vibrations of plates and shells, to the influence of viscosity on wave motion, to the phenomena of jets, and to the facts and theories of audition. This edition, taken together with the great treatise of Helmholtz on *The Sensations of Tone*, may be regarded as giving a fairly complete view of the acoustical sciences of a generation ago.

Since the publication of these classical works many briefer theoretical treatises and articles in encyclopaedias, in various languages, have appeared; these have been less condensed or less profound than the classical works, but have for the most part been based upon them. I need refer here to only one of these, the indispensable *Dynamical Theory of Sound* of Prof. H. Lamb, which, like Rayleigh's *Theory*, will endure for years because it rests so firmly on unimpugnable mechanical principles. But all of these books carry the subject along in the traditional manner and show little indication of the change in treatment that has transformed acoustics within the past decade.

During the past ten years acoustics has increased its bulk and scope enormously in consequence of the development of new methods of investigation. On the one hand the range of the frequencies with which the subject was formerly concerned

has extended into regions beyond audition—into the supersonic region and also into that of very low frequency air pulses. On the other hand, acoustics has added to itself a new area, called electro-acoustics, which is based on the telephone and its lineal descendant the loud speaker. The intimate connection with electrical science which has been brought about by the study of the telephone has reacted upon the older and more formal science of acoustics by adding a new technique, new ideas and a new jargon for expressing these new things.

The advent of these novelties has been expedited, perhaps even provoked, by the attention given to acoustics during the war; but perhaps the chief single cause of the practical importance now accorded to acoustics is the rise of broadcasting. This new popular diversion has greatly extended and altered the study of acoustics; it has jerked this study off its accustomed academic paths on to so-called practical lines. That is to say, the study of acoustics is now most keenly pursued, not for its intrinsic scientific interest alone, but also because of its utilisable results. Though this may be regretted by some, it is a fact that ought to be taken note of; and my main purpose in addressing you this evening is to draw your attention to this change of attitude and to suggest that we of the Physical Society ought to give earnest attention to these new tendencies as affecting the future development of this branch of physics.

Looking at the experimental side first, we find that many changes in the subject arise principally from the introduction of new apparatus. Most of this apparatus is electrical; and on account of this fact we have arrived at a state of things in which nearly all quantitative acoustic measurements are made after the sounds dealt with have been converted into electrical currents by means of the instrument named—very badly—a telephone transmitter or microphone. This mode of conversion has been familiar for more than fifty years; but not until recent times have types of microphone been developed which could translate sound into currents without great distortion of some kind or other.

Of these pieces of apparatus the refined condenser microphone, which depends upon the variation of electrical capacity caused by air pressures deflecting one plate of the condenser, is the most noteworthy instrument. But it would have been almost useless without that modern universal instrumentality, the triode amplifier. The modern amplifier is a marvellous tool. It can be constructed so as to magnify electrical currents and voltages conveying powers less than 10^{-10} watt. It can, if desired, amplify these minute fragments of electrical power, say 10^{11} times; and amplifiers can be designed to accomplish such feats at frequencies below the audible range and above the range as well as within it. Evidently, the combination of a sensitive microphone and a high amplifier solves the problem of converting even a feeble pressure variation in air into a measurable electrical current of equal frequency. As a fact it is now possible to detect variations of air pressure of acoustic frequency which are too faint to affect the most sensitive ear.

Not only detection but also production of air waves has been made easy by application of electrical science to experimental acoustics. It has long been possible to convert electrical currents of acoustic frequency into sound waves by aid of a

so-called telephone receiver. Of recent years it has become possible to produce electrical currents of any desired frequency and to vary the frequency easily. By means of electric amplifiers such currents can be magnified before being converted into sound by the telephone receiver. Thus it is possible now to produce sound waves which are smoothly variable in frequency and in intensity through and beyond the range of human audition. In contrast it is possible also to produce, by aid of a tuning fork or quartz crystal maintained in vibration by a triode, sounds of great constancy in frequency and intensity. All these aids to research and practice constitute another triumph of the triode which is too well known to describe here.

In this category of the aids that electrical apparatus have introduced into experimental acoustics we must not forget to include electrical filter circuits. These can be designed to pass currents higher than a given frequency or lower than that frequency, or currents between or outside two frequencies. Once sounds are converted to currents this apparatus can be employed to sift out and purify sounds of various frequencies. The power and scope of this addition to the facilities of the acoustic laboratory have to be experienced to be appreciated fully.

These are the main pieces of apparatus that electrical science has thrust into acoustics. But, in fact, when a sound is converted into electrical current, the whole equipment of the electrical laboratory, from the Wheatstone bridge to the coherer, becomes available for measuring. Nearly all the modern additions to acoustic knowledge are implicitly indebted to this fact.

Besides these tangible adjuncts to the technique of experimental acoustics, electrical science has brought subtle assistance to the more theoretical aspects of the subject. This comes about because vibration phenomena of all kinds approximately satisfy the same linear differential equations. Inasmuch as the study of electrical vibrations in well-defined electrical circuits is easier and has been more cultivated (for practical purposes) than that of air vibrations, acoustic science profits from electrical by a free exchange of ideas about vibrations. Many acoustical problems can be translated into problems concerning electrical networks, and as there exists a great body of knowledge of such networks, the problem is often solved in the act of translation. Further, by adopting the phraseology of the electrician into acoustics, so that translation of the acoustic problem into the electrical problem becomes automatic, a language for thinking and talking becomes available and is found to clear the mind and assist reasoning. Moreover, where some new conception occurs in electrical vibration theory, we can immediately seek the analogy in sound. For instance, Van der Pol's recent discussion of the mode of vibration which he calls a relaxation oscillation and which is exhibited by the flashing neon tube in a resistance-capacity circuit translates into and explains certain features of such acoustic phenomena as the squeak caused by rubbing a slate pencil up a slate (beloved of schoolboys when slates were slates). Some of these advantages are perhaps despised by the acoustic purist, who thinks that the introduction of these electrical analogies into acoustics is merely for the purpose of supporting the electrical expert in an unaccustomed terrain; but as a fact, electrical analogy has been used by the greatest of the pure acousticians, such as Rayleigh and Lamb, and therefore the

gibe need not be feared or regarded. Rayleigh, it was, I believe, who introduced the conception of the acoustical conductivity of an orifice in a rigid barrier.

I propose to summarise rapidly some of the outstanding advances that have accrued in consequence of the new attitude towards acoustics and the new experimental technique. First I will glance at the subject of architectural acoustics—the properties of sound waves in an auditorium. This study has been greatly assisted by the ability to produce pure monotone sounds by passing a filtered alternating current into a loud speaker; and by being able to employ nearly distortionless microphones and amplifiers for picking up sounds of all ordinary intensities and pitches. The fundamental conception, first stated I think by W. C. Sabine, is that a sound created in a closed room is ultimately distributed by repeated irregular reflections so that (leaving foci out of account) each cubic centimetre of air is traversed by waves proceeding in every direction. He clearly stated that sound can only decay by absorption in the walls and furniture and he showed that in general there were several hundred reflections before the echoes die. It has long been known that when the decay is unduly slow the acoustics of a room are defective and the defect is now called reverberation. Sabine introduced the idea of measuring the reverberation time of a room, and defined the duration of reverberation as the time taken for a sound of a million times that of bare audibility to die down to its threshold value. This definition was given, and Sabine made measurements of the reverberation times of rooms, before modern electric methods came to the aid of acoustics—but the subject attracted little attention. Since the arrival of the amplifier, however, the subject has sprung to life in every country. Accuracy has been introduced into our prescriptions of the best reverberation value for speech, for singing, for instrumental music, respectively, and in halls of various sizes. The acoustic data of some of the principal concert halls and debating chambers of the world have been determined and published; empirical laws connecting the optimum reverberation values with the volume of these halls have been established; and the absorption of materials and of objects placed in rooms has been determined by methods based on Sabine's definition of reverberation. As a consequence it is now possible to measure the defects of a given auditorium and to calculate with fair certainty the amount of absorbing material that should be permanently introduced to correct these defects. Better still, it is now easy to plan buildings before their erection to have a desired reverberation time—in fact to be of good acoustic quality; and the design of noise-proof houses and laboratories has progressed very rapidly.

As another example of the range and power of the new methods, take the growth of psychological acoustics. Firstly, we now have firm data of the intensity range of the ear. At a frequency of 2000 cycles the minimum audible sound, according to Fletcher, corresponds with a pressure of 6×10^{-4} dynes per sq. cm. and an amplitude of 2×10^{-9} cm.; while pressures up to 6000 dynes per sq. cm. and amplitudes up to $1{\cdot}7 \times 10^{-4}$ cm. are still recognisable as sound. Changes of intensity of a monotone, we now know, cannot be perceived under most favourable conditions unless they exceed 10 per cent. of the original; and thus the total audible range of intensity can be traversed in about 270 barely perceptible steps. Taking comfortably

perceptible differences of intensity, it has been suggested that 120 equal percentage steps may be arbitrarily assumed to cover the range from faintest to loudest sounds. Thus 120 degrees of a sensation scale have been adopted in some quarters as corresponding to the whole range of the ear; and consequently if one sound contains x times the energy of another, the sensation ratio is $10 \log_{10} x$ degrees. In Sabine's method of reverberation, if we take threshold audibility as zero his standard initial intensity of the test sound is 60 degrees higher on the sensation scale.

I must refer here to the other method of approach to these new ideas relating to the measurement of the sensation of sound. It has long been obvious that we may measure either the energy flow that causes the sensation, or, alternatively, the strength of the sensation. The measurement of this strength is rightly a psychological problem, and the main suppositions in attempting such measurement are that changes of sensation can be compared with some accuracy and that our perceptions of such changes are related to the energy changes by the rule that equal steps of sensation correspond with equal percentage rises of the power causing the sensation; a geometrical progression in power corresponds to an arithmetical progression in sensation. In other words the logarithms of the steps in power may be used as the measure of steps in the corresponding sensations. This recalls the mode of expressing stellar magnitudes.

Telephone engineers for some years have been using a very convenient standard of change of aural sensation called the 800 cycle standard mile. It is the step in sound sensation which is perceived by an observer when he compares two equal telephones, one at the input end and the other at the output end of a mile of standard cable, the frequency being 800 cycles—the difference of strength of sound being due to electrical dissipation in the (artificial) cable. This step in sensation corresponds approximately to a step of 25 per cent. in power. Hence has arisen a proposal to standardise a power step of ratio $10^{0\cdot1}$ (which is equal to 1·259) and to generalise it so as to apply it to express the ratio of two powers of any nature and any frequency. The corresponding standard step in sensation is called a "transmission unit," an unfortunate misnomer. It can be calculated from any given set of measurements by taking the logarithm of the power ratio to the base 1·259, or, what is the same thing, taking the logarithm of the power ratio to the base ten and multiplying by ten. We therefore have the following relations between the ratio of the powers causing sound sensations and the number of standard steps of sensation by which the sensations differ.

Power ratio	1·259	2	4	10	100	1000
Steps in sound sensation or "transmission units"	1	3	6	10	20	30

It is interesting to note that that other important subject-matter in psychological acoustics, the sensation of pitch, proceeds by ratios; and, moreover, that the makers of keyboard instruments have adopted, without knowing anything of logarithms, a logarithmic scale for their instruments—equal distances along the keyboard represent equal ratios in pitch. Consequently we conclude that the most human

way of representing acoustical response relations graphically is to use transmission units for intensity as one coordinate and logarithms of frequencies as the other coordinate.

In the acoustics of the ear one of the most remarkable applications of the new technique is that of Wegel and Lane in a research which carries to a well-defined conclusion the observation of A. M. Mayer in 1876 that "no sound, even when very intense, can diminish or obliterate the sensation of a concurrent sound which is lower in pitch." This qualitative observation has now been made quantitative, and though it proves to be a much more complicated matter than Mayer thought, it is broadly confirmed. Low tones are easily masked if not much lower than the masking tone (except in so far as beats may enter) and are not easily masked if much lower than the masking tone; and high tones can be masked by low tones when the ratio of the intensity of the latter to the former reaches a definite value. Wegel and Lane have separated the type of masking which occurs in the ears from that which occurs in the brain. The same investigators have applied their methods to the successful unravelling of old problems concerning combination tones arising from the non-linear response of the ear.

The advances just described in our knowledge of the acoustics of buildings and of the mode of operation of our ears, have led to or aided many noteworthy practical applications. The most important of these is, possibly, that of magnifying the voice of an orator so that it may be made audible to large audiences. The equipment is simple in appearance—a microphone, an amplifier, and a loud speaker—but its development to its present nearly perfect condition has involved much labour. In rooms and halls the problems of echoes and foci, of resonances and reverberation, all tend to be exaggerated by the magnified intensity of the sounds; and electrical vibration of the apparatus caused by waves from the loud speaker reaching the microphone (in addition to the original waves) easily leads either to singing or to distortion. In all cases the demons of distortion obtrude themselves at every stage of the acoustic and electrical processes occurring within the chain of apparatus. The most frequent type of distortion has proved to be that caused by the unequal treatment of sounds of different frequencies, in either the microphone, the amplifier or the loud speaker; but in the study of these public address apparatuses still other types of distortion, originally unsuspected, have appeared. One of these arises from the fact that when a sound is started or stopped a transient vibration of the chain of apparatus, or of the auditorium, may be excited, and this adds itself to the original sound; another trouble may arise from the asymmetric treatment, in any portion of the apparatus but especially in the amplifier, of pure sounds—this introduces sum and difference tones which may cause dissonances in music or unintelligibility in speech. Inasmuch as it is usually found necessary to amplify the wanted sounds until they are ten thousand times as intense as the unwanted disturbing sounds arising from a big audience, the distortions, dissonances and transients just mentioned have placed many difficulties in the path of the development of these appliances for assisting orators. Fortunately the ear is so bad a judge of intensity of sound that exact equality in the magnification of each of the frequen-

cies in a complex sound is not required. In fact certain frequencies of the speech range may be magnified in intensity ten times as much as certain other frequencies in that range, without the ear noticing the inequality of amplification; and certain frequencies may be altogether omitted by the apparatus from the reproduction of a speaker's voice without loss of intelligibility and naturalness. As a consequence of these inherent tolerances in the human ear it is now possible to obtain apparatus capable of helping a public speaker to address great audiences without any suspicion of artificial aid; and apparatus has been developed competent to spread intelligible speech to an audience of a million persons. It has, however, been found better to attempt such feats by dividing the amplified speech currents among a number of well-distributed loud speakers than to deliver it all to one powerful loud speaker; for it is a curious fact that when the intensity of a complicated sound—either speech or orchestral music—is greatly magnified perfectly equally at every frequency—that is to say without any distortion whatever—and delivered to the ear, the listener believes that the sound is badly distorted, partly as a consequence of the non-linearity of the mechanism of the ear.

Just to indicate the closeness of the approach to perfection which these modern applications have reached I might mention that speech sounds ranging from 100 to 6000 in frequency can be magnified fairly equably; and most of the spectrum of the fricative consonants *s*, *f*, *th* between 4000 and 10,000 in frequency can be reproduced. The intensities of the sounds to be dealt with vary in the ratio 1500 to 1 in the case of speech, and in the ratio 100,000 to 1 in the case of music.

And now we must turn to the developments that have taken place in acoustics as a consequence, at least in part, of the introduction of electrical analogy into the theory of sound vibrations. The analogy is possible because the differential equations relating to the transmission of electrical oscillations through electrical networks are linear like those relating to the propagation of sound vibrations. The analogy has long been obvious to the mathematician and may prove of little assistance to him; he may say, justly, that it does nothing other than alter the nomenclature; but even so we may admit that the use of the analogy has helped many recent investigations.

Electrical ideas began to enter acoustics explicitly when the telephone receiver, regarded as an element of an electrical circuit, came to be investigated scientifically. The mathematical aspects of this investigation are very similar to those made familiar to physicists by the study of the vibration galvanometer and various types of electrical machinery such as the synchronous motor. In all cases we have a winding carrying alternating current, a magnetic field acting mechanically on the winding, and relative motion between the two. And when the rotor of a motor or the diaphragm of a telephone is maintained in steady motion by the electro-dynamic forces due to the alternating current, the motion induces an E.M.F. which is proportional to the velocity and may therefore be regarded as due to a special impedance in the windings. That is to say, in the case of the telephone, the inertia, the elastic quality and the energy dissipating property of a diaphragm can, at any given frequency, be represented quantitatively by a hypothetical reactance and

resistance provided the mechanical and electrical units are properly coordinated. This conception is implicit in some of the remarks of Rayleigh and of M. Wien (who, in 1891 equipped a telephone diaphragm with an optical lever and so made the first vibration galvanometer) but it was not until 1912 that it was formulated adequately by Kennelly and Pierce and exhibited in the so-called "vector" diagrams or crank diagrams of the electrician, and given the name of "motional impedance." The motional impedance included, of course, terms or components covering the production by the diaphragm of sound waves or other motions in air. The conception was quickly carried into other branches of acoustics, for instance by Hahnemann and Hecht, and by Webster.

Webster, especially, emphasised in 1914 the use of the conception of acoustic impedance for replacing by a single complex number all the quantities involved in the reaction of an acoustical system. He proceeded to write down formulae for the typical impedances most frequently met with in pure acoustics, namely those due to inertia, elasticity and the production of free waves. He then obtained general expressions for the impedance of the open end of cylindrical and tapering tubes, and incidentally completed, by comparatively easy steps, Rayleigh's early discussion on the properties of horns. His solutions for the conical and the exponential horns are now well known and have been elaborated by a number of followers.

As a consequence mainly of this work of Webster's our knowledge of the operation of horns has been much clarified. At the narrow end of a horn the energy enters in the form of plane waves and so the source works at maximum efficiency; at the open end the reactance due to inertia is minimised and the rate of working on the atmosphere increased by the largeness of the area. Consequently the amount of sound emitted is increased, the waves are little reflected at the open end, especially if their length is small compared with the diameter, and standing waves or resonances are minimised.

What happens to sound waves passing between the narrow end and the open end of a horn depends upon the form of the tubular surface. This has been thoroughly worked out by the methods of Webster and it is well known that the exponential horn, that is, the horn of which the sectional area increases by equal percentages in equal distances, is much better than the conical horn in certain respects. For instance, quoting Flanders, the impedance "looking out" of a hole 0·7 inch diameter in an infinite plane wall, which is variable with frequency, reaches 80 per cent. of its ultimate value at a frequency of 9300 cycles; a certain conical horn applied to the hole brings this 80 per cent. value at 4200 cycles; and a comparable exponential horn brings it at 250 cycles. The horns do not alter the ultimate impedance of the hole in the wall. As a consequence we may treat a horn as having an impedance equal to that of an infinite cylindrical tube of diameter equal to that of the small end of the horn for air vibrations of frequency above a certain lower "cut off" limit which is determined by the geometrical figure and dimensions of the horn. The impedance of an infinite cylindrical tube is, of course, a pure resistance, as energy passes along it from the source undisturbed by reflections from the end.

These new methods of viewing acoustic problems have had many interesting applications. One of these may be briefly described since it has resulted in great improvements in the rendering of music and speech by the gramophone. As is well known the vibrating parts of a gramophone comprise the needle, a lever (called the stylus lever) which rocks on a fulcrum, a diaphragm which forms the back of the sound box, the air chamber of the sound box, the tone arm and the horn. The three first-named items possess inertia and pliancy and therefore contribute to the equations of motion of the system terms analogous to those introduced by inductance and capacitance into electrical equations; the other items contribute only a resistance. To the electrical expert the easiest way of utilising this analogy is to represent the inertia of each part by a suitable inductance coil, its pliancy by a condenser, and its power of consuming energy without phase change by a resistance. The chain of mechanical vibrating elements is thus replaced by a chain of inductances and condensers terminated by a resistance; such an electrical chain constitutes the "artificial line" or the "filter circuit" of the telephone engineer. The properties of these electrical chains have been thoroughly worked out and all the mathematical requirements for ensuring the easy passage of vibrations from element to element without reflection at the junctions are known. Hence can be deduced immediately the appropriate mass and pliancy of the stylus arm, of the diaphragm and of the sound chamber, for the efficient cooperation over a large range of frequencies between these elements and the energy-consuming "load" imposed by the tone arm and horn. The result has been a notable improvement, in both output and quality, of the gramophone, and the instrument is now worthy to deal with the modern electrically-made records.

From this rapid review of the change that has come over the science of sound since the publication of Rayleigh's treatise fifty years ago, it will be seen that the new acoustics is Baconian, that is to say, it is being prosecuted with a view to rendering services to mankind rather than from the motive of scientific curiosity. It is, I think, not being studied so eagerly in this country as in other large countries, notably America and Germany; and I suggest to the Physical Society that this is unfortunate. If this address has the effect of bringing it to the attention of physicists and of prompting those who control college courses and examinations to pay attention to the possibility of encouraging its systematic study, my purpose will have been fulfilled.

THE DEPENDENCE OF THE MOBILITY OF IONS IN AIR ON THE RELATIVE HUMIDITY

By EZER GRIFFITHS, D.Sc., F.R.S. and
J. H. AWBERY, B.A., B.Sc.

Physics Department, National Physical Laboratory, Teddington, Middlesex

Received February 19, 1929. *Paper contributed to the discussion held on March* 8, 1929

ABSTRACT. The relationship between relative humidity and the mobility of negative ions in air at atmospheric pressure has been examined with a view to possible applications in hygrometric measurements. The method employed was a modification of Zeleny's original method, the end of a wind channel being closed by a disc of gauze fitted with a guard ring through which a steady stream of air of definite humidity was pumped.

The motion of the negative ions due to the action of the air stream was balanced by a counter potential gradient, and the mobility was deduced from the critical potential required to produce a balance. The rate of air flow was measured by means of a Ewing ball and tube flowmeter, a hollow glass sphere being used to make it suitable for low air rates. Efforts were made to construct a direct indicating instrument by amplifying the current sufficiently to obtain deflections on a robust moving-coil indicator, but the variations in the amplification factor were found to be too serious.

---

§ 1. INTRODUCTION

When time is not an important factor humidity can be measured by the customary methods without the use of elaborate apparatus. It will, however, be recognised that humidity measurements, like most other physical measurements, if rapid observations have to be made, call for apparatus differing radically from that which may be quite suitable if a certain amount of time lag in the instrument is of no consequence. It was desirable, in connection with some special investigations, to have the means of following fairly rapid changes of humidity, and it was suggested that the variation with the humidity of the air of the mobility of the ions produced from some such source as a hot wire might constitute the basis of a method of measuring the humidity. As no information was available at the time as to the relation between mobility and relative humidity, we decided to investigate this effect as a preliminary step in the development of the method.

Zeleny* had shown that the mobility of the negative ion in dry air was 1·24 times greater than in saturated air, and when the work described in this paper was practically complete Tyndall and Grindley† published their work on the relationship between mobility and the moisture content of the air, using a fundamentally different method.

* *Proc. Roy. Soc.* A, **46**, 238 (1900).

† *Proc. Roy. Soc.* A, **110**, 341 (1926).

§ 2. PRELIMINARY EXPERIMENTS

In selecting a method of measurement we had in view the desirability of being able to adapt it later as the basis of the design of an instrument for measuring humidity. The first plan tried was to use a hot wire as the source of ions and observe the variation of the ionisation current between it and a coaxial cylinder when the applied potential was less than that required to produce saturation. The difficulty we encountered with this method was the disturbance due to convection currents from the heated wire. Attempts to use a wire coated with calcium salts so as to obtain an adequate supply of ions at a lower temperature were not successful.

Our attention was then directed to the possibilities of radioactive material. A thin layer of uranium oxide was spread over an aluminium plate and covered with layers of thin aluminium foil, the number of layers being so adjusted that the α-rays just emerged and produced an ionised layer of gas close to the plate surface. A similar plate was supported parallel to the one covered with uranium oxide, the former being connected to an electrometer. The current passing between the plates for various applied potentials was determined. Owing to the weak ionisation of the uranium salt after filtering through the aluminium foil, saturation currents were obtained with potential differences as low as 10 volts.

While there was no difficulty in measuring the small ionisation currents produced, a very formidable difficulty was encountered in the variable contact-potential effects. With no applied voltage a small ionisation current which varied erratically was obtained. Changes of temperature had a marked influence on the value of the contact-potential, and ultimately this simple method of experiment had to be abandoned in favour of the one described below.

§ 3. WIND CHANNEL METHOD

The principle of the method is to balance the movement of the ions under the electric field by an opposing air blast. In the application of this idea to the design of an apparatus for measuring mobility when moisture is present in the air stream, two sources of error have to be guarded against:

(1) Electrical leakage over the surface of insulators due to a film of absorbed moisture has to be prevented.

(2) Uniformity in velocity distribution across the air stream has to be obtained.

Surface leakage can be reduced to a negligible magnitude if the insulator be kept warm so as to prevent moisture condensation. In these experiments a sleeve was slipped on to the ambroid insulator, and this was heated by a small resistance coil wound round it.

The question of the distribution of the velocity across the air stream was one which received much consideration.

Zeleny in 1898 tried the air stream method. He employed a current of gas flowing through a tube, the end of which was fitted with a disc of gauze which very nearly fitted it. He abandoned the method in favour of one in which the electric field on the ions was perpendicular to the air blast. Now it is a well-known fact

that the velocity of the stream in the vicinity of the walls of a tube falls off to zero at the boundary. Consequently, the air blast method fails if an attempt is made to work with the entire stream which flows through the tube, since there is no opposing blast to balance the electric field in the region of the walls.

By making one modification in Zeleny's original method we were able to adapt it for the determination of the variation in mobility with humidity. Instead of using a single piece of gauze to close the end of the tube, we divided the gauze into a central disc and a guard ring, see Fig. 1. The central disc was made to fit very accurately into the hole in the gauze, but was isolated electrically. (In Fig. 1 the gap between central disc and guard ring has been exaggerated for the sake of clearness.)

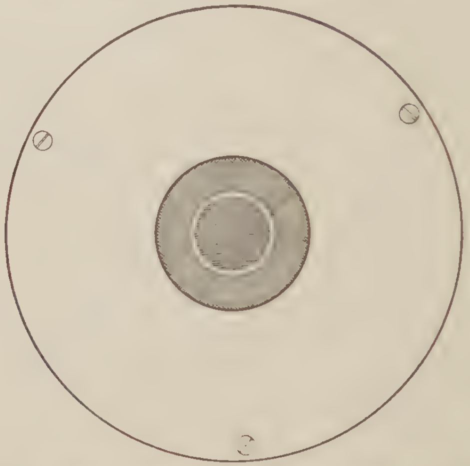


Fig. 1. Gauze electrode and guard ring at end of tube.

The general arrangement of the apparatus employed in this investigation is shown in Fig. 2. A small air blower delivers a steady stream of gas through three wash bottles in series, containing sulphuric acid solution of definite concentration. The gas then passes through a glass wool filter on to a flowmeter. From the flowmeter the gas passed into a metal tube 2 inches in diameter by 5 feet in length. In one end of this tube a pad made of metal ribbon was inserted and connected to earth. The function of this metal filter was to remove any ions formed by the bubbling of the air through the sulphuric acid solution.

The electrometer and earthing key will be seen in the bottom right hand corner of Fig. 2.

§ 4. DESCRIPTION OF FLOWMETER

This flowmeter is of the Ewing ball and conical tube* type, but instead of the usual solid steel ball we employed a hollow glass sphere so as to adapt the instrument to the measurement of low rates of gas flow.

* Sir J. A. Ewing, "A Ball-and-Tube Flowmeter," *Proc. Roy. Soc. Edin.* **45**, 308 (1925); J. H. Awbery and Ezer Griffiths, "Further Experiments with the Ewing Flowmeter," *op. cit.* **47**, 1 (1927); *Proc. Phys. Soc.* **39**, 296 (1927).

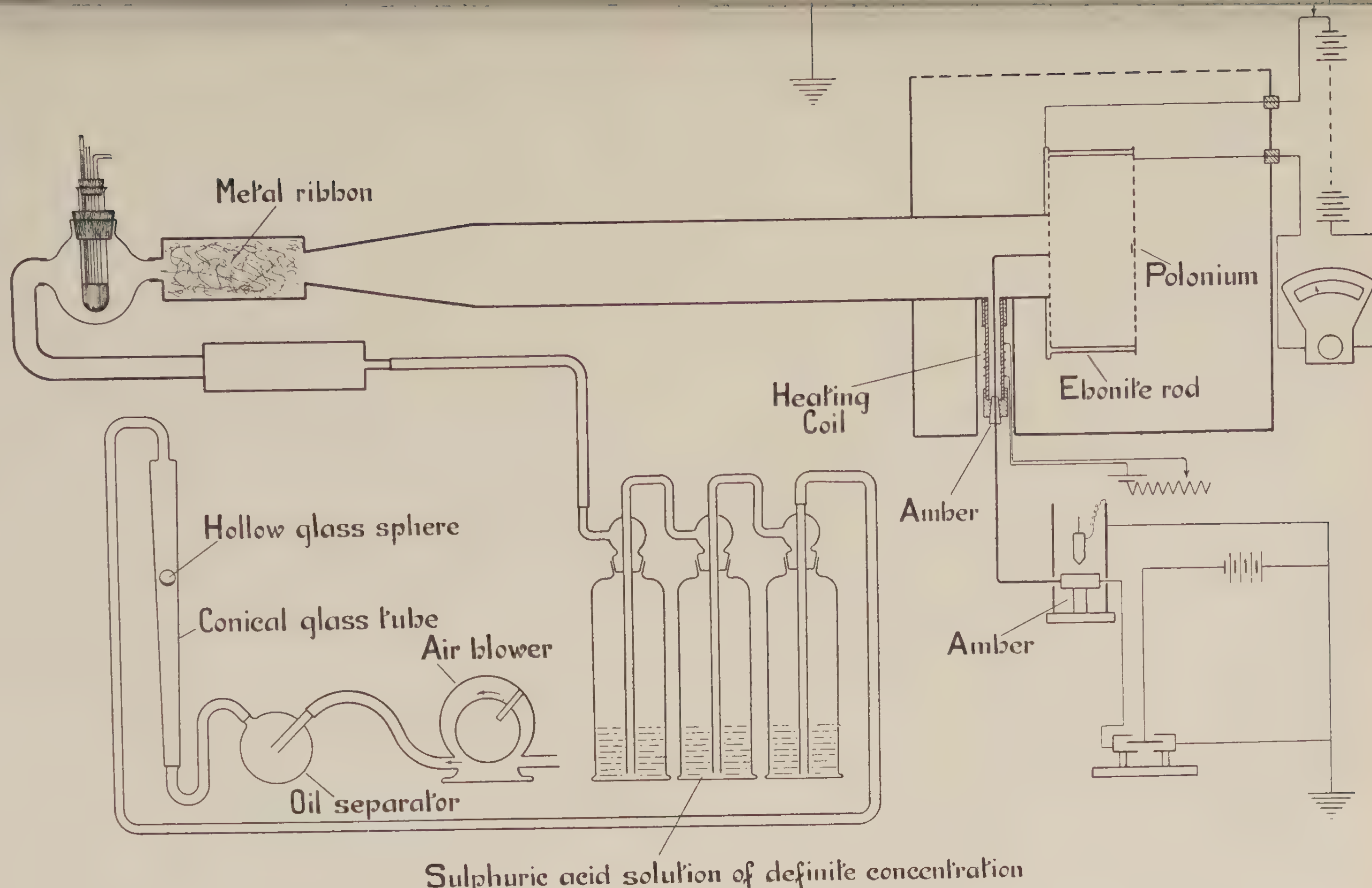


Fig. 2. General arrangement of the apparatus.

A curious fact came to light in connection with this flowmeter. With the ordinary form of the instrument, when using a solid sphere, it is found that with the tube set vertically the position the ball takes up in the tube for a given rate of flow depends on its mode of vibration transverse to the axis of the tube. When the ball is chattering violently, it is carried to a higher position in the tube than when it is quiescent. This characteristic necessitates using the Ewing ball and tube instrument with the tube inclined to the vertical. The present instrument with the hollow sphere was found, on the other hand, to behave quite satisfactorily with the tube set vertically, and this is to be ascribed to the fact that the sphere is not perfectly symmetrical in form. The walls of the sphere were non-uniform in thickness, and it was observed that when the instrument was functioning the sphere would set itself with the thick lenticular segment lowermost. In this position it rides quite steadily and remains in the axis of the tube with a clear space all around for the air to flow past.

By independent tests it was proved that, to the order of accuracy required, the calibration of the flowmeter did not depend on the relative humidity of the air stream.

§ 5. METHOD OF OBSERVING

In the experiments the current of gas was maintained constant throughout and the strength of the electric field was varied. Constancy of the gas stream was secured by an observer watching the flowmeter and adjusting the speed of the blower so as to keep the glass sphere floating between narrow limits.

The ionisation was produced by a small quantity of polonium carried on a sheet of gauze, as shown in Fig. 2. Various negative potentials were applied to this gauze and the current flowing across to the disc in the end of the tube was measured by timing the transits of the electrometer spot in the usual manner. The distance between the disc of gauze and the polonium was 10 cm. The disc of gauze was connected to one pair of quadrants of an electrometer through an earthing key of the design shown in Fig. 3. The other pair of quadrants, the guard ring of gauze and the flow tube were all maintained at earth potential. Considerable difficulty was encountered with earthing keys of various types until the form shown in Fig. 3 was evolved, which has proved very satisfactory. This consists of a platinum-faced disc of copper carried on an amber pillar; the disc was connected to one pair of quadrants. A heavy plummet of copper with a solid sphere of platinum set into the end was supported on a thin string above the disc. The plummet was connected to earth. By a lever operated from a distance the plummet could be raised so as to break contact with the face of the disc, thus allowing one pair of quadrants to rise in potential owing to the charge carried by the ions reaching the gauze.

§ 6. TYPICAL SERIES OF OBSERVATIONS

It is unnecessary to burden the paper with the whole of the experimental data, and the general character of the results will be seen from a study of Fig. 4 which shows a few typical series of observations. The straight lines give the relation between

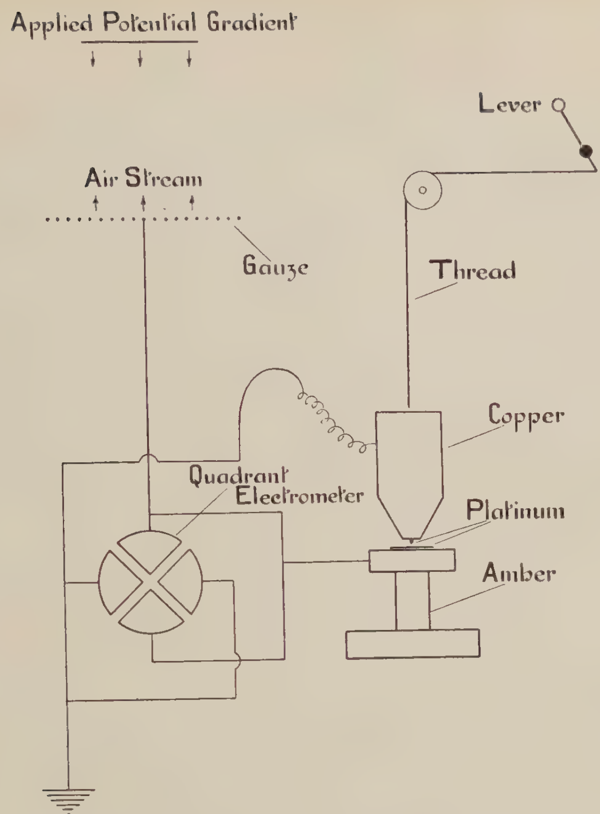


Fig. 3. Connections of electrometer and earthing key.

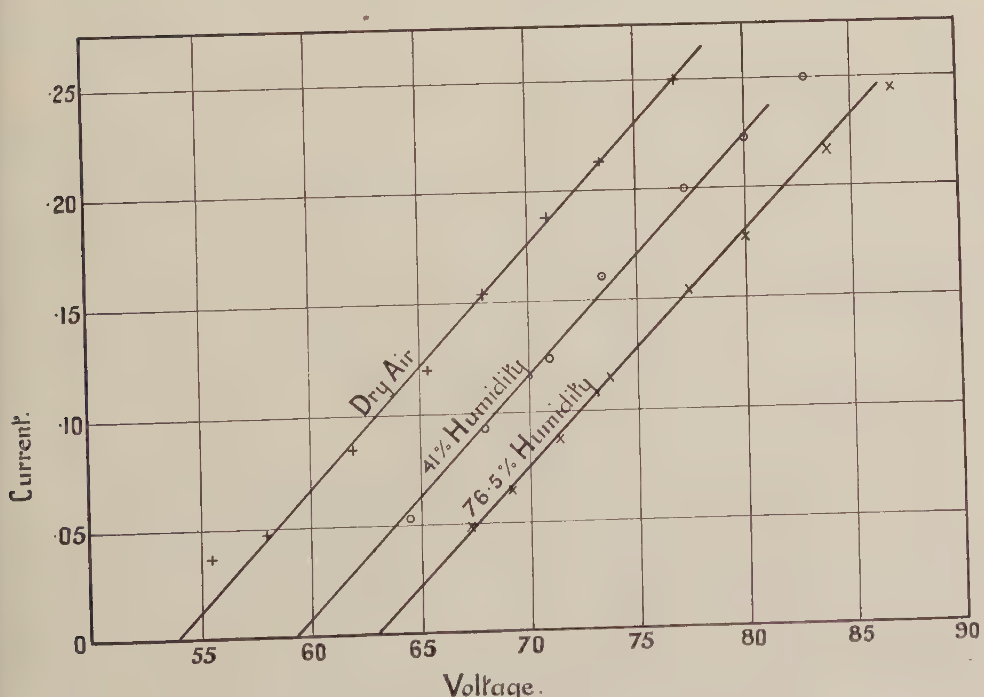


Fig. 4. Three representative sets of observations.

the applied voltage and the current in three separate experiments, using air of different humidities.

The actual balancing electric field was not observed directly, but the intercept on the abscissa was taken as the value of the potential which would set up just sufficient velocity in the ion to balance the velocity due to the air stream.

§ 7. RELATION BETWEEN MOBILITY AND RELATIVE HUMIDITY

The results obtained in the experiment are summarised in the graph, Fig. 5. The values are expressed in terms of the fractional increment of potential difference required to maintain the same velocity (taking as the initial value the potential required for air which has been thoroughly dried by being passed through sulphuric acid).

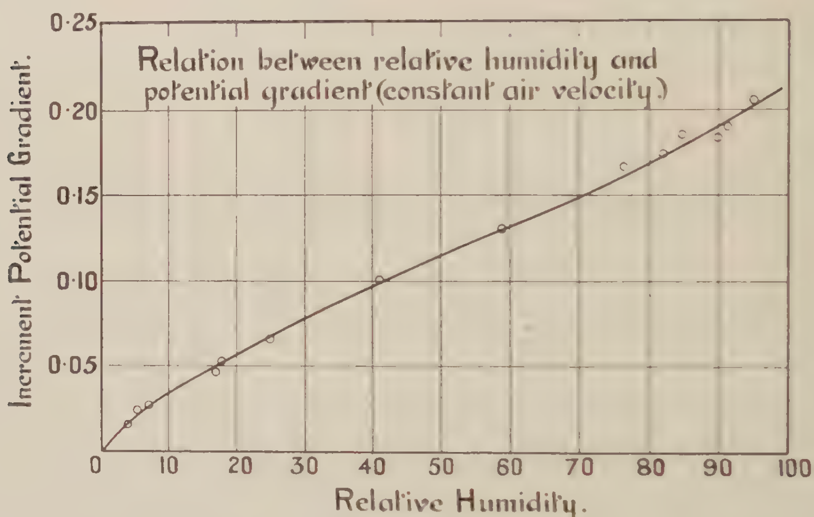


Fig. 5. Variation of the mobility of the negative ion with the relative humidity of the air in which it is moving.

In Table 1 the values taken from Fig. 5 have been converted to relative mobility, taking the value for the ion in dry air as unity.

Table 1

Relative humidity (%)	Mobility relative to mobility in dry air
0	1·00
20	0·95
40	0·91
60	0·88
80	0·86
100	0·82

§ 8. APPLICATION OF THE EFFECT OF HUMIDITY ON MOBILITY TO THE DESIGN OF A HYGROMETER

For the purposes of practical hygrometry, the use of a quadrant electrometer as a current-measuring instrument is not feasible. Many attempts have been made to devise amplification devices so as to magnify the minute currents to a sufficient extent to render possible their detection by means of moving coil indicators. A valve amplification system has been tried in this connection, and it was found possible to obtain deflections of the order of 6 cm. on the scale of a microammeter of the pointer type for a change of humidity of 100 per cent. The difficulty encountered with this arrangement was that the amplification factor was not constant and that sufficient magnification could not be obtained to permit of the use of a more robust indicator.

Further developments in this method of hygrometry would be feasible if a more powerful source of ionisation were available, for when the ionisation is feeble surface leakage becomes a limiting factor.

§ 9. ACKNOWLEDGMENTS

The authors desire to express their thanks to Sir Ernest Rutherford, O.M., P.R.S., for the suggestions which led to the investigation: to Dr G. W. C. Kaye for advice on the work: and to Mr A. Snow, Observer in the Physics Department, for his resourcefulness and skilful assistance.

DISCUSSION

Dr J. H. Vincent. I should like to suggest the importance of carrying the experiments shown in Fig. 4 to higher values of field. The apparatus seems to be a very convenient tool for investigating many of the still unsolved problems on ions in gases at atmospheric pressures.

Mr K. L. Goodall (communicated): I should like to ask the authors whether in their search, mentioned in § 8, for an ionisation-current amplification device to replace their quadrant electrometer they have tried the "compensated" valve amplifier bridge developed by Mr C. E. Wynn-Williams?* Some experience with this apparatus suggests that it should be suitable for the purpose in view, as it possesses a steady galvanometer zero and a fairly constant amplification factor of about 10^5 when power valves are used, permitting the use of a robust low sensitivity galvanometer to measure the amplified ionisation currents.

Reply by Dr Ezer Griffiths: In reply to Dr Vincent, I quite agree that it would be of interest to study the form of the curves in Fig. 4 with higher values of the potential gradient. The apparatus described in the paper would probably be quite suitable, but I doubt whether we shall have an opportunity of studying it ourselves. We thank Mr Goodall for his suggestion.

* *Proc. Camb. Phil. Soc.* **23**, 811 (1927).

SOME PROBLEMS RELATING TO THE MOBILITY OF GASEOUS IONS

By Prof. A. M. TYNDALL, D.Sc., F.Inst.P.
With a note by C. F. POWELL

Address contributed to the discussion held on March 8, 1929

ABSTRACT. The address deals with the following topics: 1. Established results. 2. The difference between positive and negative mobilities. 3. The effect of vapours. 4. Mobility in pure gases. 5. Positive ions of short age. 6. Suggestions as to future progress. *Note by Mr C. F. Powell*: An apparatus of the "four gauze" type has been designed for experiments with highly purified gases.

§ 1. INTRODUCTION

Some of the main experimental facts relating to mobility are now from 25 to 30 years old. Yet to-day there is no satisfactory quantitative theory of the motion of an ion through a gaseous medium. Moreover, in recent years some results have been obtained which not only have received no satisfactory explanation, but also, taken at their face value, appear to be mutually inconsistent.

If we omit from consideration for the moment (*a*) ions in highly purified gases, (*b*) ions of short age, we may number the following results amongst those which are well established:

(1) The product of mobility and pressure is a constant over a fairly wide range of pressure.

(2) The mobility of a negative ion in most gases as distinct from vapours is greater than that of a positive ion.

(3) The mobility of a heavy ion such as "radioactive recoil atoms" in air is the same as that of ordinary air ions. This is also true for ions formed in hydrogen, carbon dioxide or various vapours, and measured in room air.

(4) The presence in the gas of a vapour with a marked dipole causes a reduction in the mobility of the negative ion and, to a lesser extent, in that of the positive ion.

Under these conditions we may accept the view that the ion is a charged core surrounded by a cluster of neutral molecules. Of the various treatments of the motion of a rigid or "static" cluster which have been given, that of Langevin, while admittedly incomplete, is probably the most satisfactory, since it is not based upon considerations of mean free path. Its consequences have been analysed by Hassé*.

A truer picture would, however, be given by regarding the cluster as "dynamic,"

* Hassé, *Phil. Mag.* 1, 139 (1926).

governed by statistical laws and of a virtual size determined by the concentration of neutral molecules in the neighbourhood of the core at any given instant. The expression for its mobility will then include a term which will be a measure of the probability that a given neutral molecule will be found in an element of volume at a distance r from the charged core. When the drift of the ion due to the applied field is small we are probably justified in applying Boltzmann's theorem by which this probability is proportional to $e^{W_r/KT}$ where W_r is the work required to separate to infinity the neutral molecule from the ion. W_r in turn will depend upon the magnitude of (i) any permanent dipole that the molecule may possess, and (ii) any induced dipole set up in the molecule by the field of the ion. The function W_r falls off with distance according to the inverse square law owing to (i) and the inverse fourth power law owing to (ii). Whichever view we take of the nature of a cluster, there does not appear to be any strong evidence for supposing that it is equivalent to a layer more than one molecule thick.

For convenience, we may call this term the clustering factor, and its value will depend upon the combined magnitudes of the permanent and induced dipoles in the molecule.

Amongst other factors which we might expect to influence the mobility are:

(*a*) The mass of the ion (M) and the molecule (m).

(*b*) The distance between the centres of a molecule and a cluster at collision (σ). For instance, in Langevin's treatment, this enters as a term proportional to $1/\sigma^2$.

(*c*) The density and temperature of the gas.

With a fully formed cluster there is reason for supposing that (*a*) is of minor importance compared with (*b*). In Langevin's formula M and m enter as a term $(1 + m/M)^{\frac{1}{2}}$ which tends to unity for large clusters. The experimental evidence suggests that it is sensibly unity. Thus, as stated above, the radioactive recoil ions and air ions have the same mobility in air. The other results under (3) (page 248) suggest that when an ion produced in one gas A enters another gas B it sheds the molecules of A and attaches to itself the molecules of B. This is to be expected from dynamical considerations. The fact, however, that it makes no difference to the mobility whether the gas A is hydrogen, a heavy vapour, or the air in which the mobility is finally measured is hardly in accordance with theory. One would expect some difference because the value of σ should vary with the nature of the core. Thus from Langevin's theory Hassé has calculated that the mobility of a hydrogen ion measured in air should be 15 per cent. higher than that of an air ion. This difference is well beyond the experimental error of the experiments. It should, however, be stated that the experiments were carried out in moist gases, i.e. in the presence of foreign polar molecules which tend to increase σ; they deserve repetition with thoroughly dry gases.

§ 2. DIFFERENCE BETWEEN POSITIVE AND NEGATIVE MOBILITIES

On the static cluster view the difference between the mobility of positive and negative ions is qualitatively explained in terms of the difference in dimensions of the core, in one case a charged molecule, and in the other an electron. With a single

layer of molecules in the cluster both M and σ are greater for positives than for negatives, and the positive mobility is the smaller of the two. Hassé has shown that in certain gases including air and CO_2, Langevin's theory may be applied in this way to give quantitative agreement with experiment, though it is impossible to fit helium and hydrogen into the same scheme.

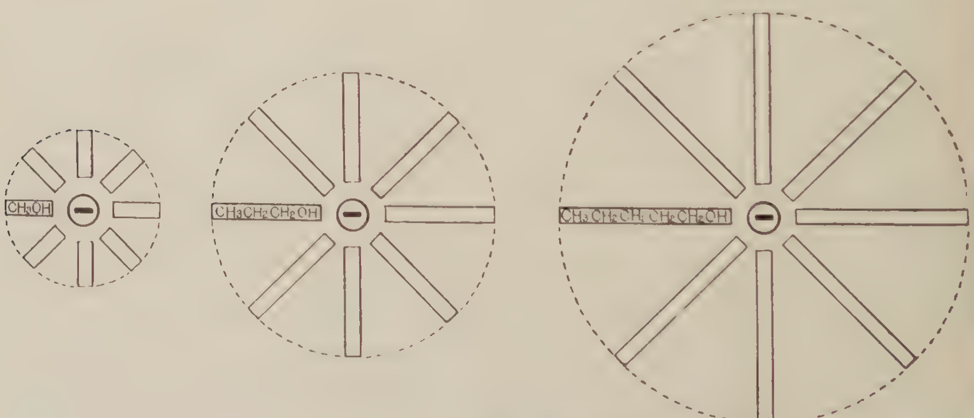
On the dynamical cluster view the lower mobility of the positive ion is not so obvious, because, owing to the fact that the minimum distance of approach to a positive molecule will be greater than that to an electron, the magnitude of the clustering factor will be reduced. Alternative views on this difference have been put forward, namely:

(i) The negative ion may start as an electron and only acquire a cluster after having travelled an appreciable fraction of its path in the measuring apparatus. In this case, its mean mobility would be higher than that for a full life as a cluster. That this is not the explanation is shown by the results obtained by Grindley and the author by a method which was designed, amongst other things, to eliminate any effect of this kind.

(ii) The negative ion is a charged molecule of which the extra electron throughout its path in the measuring apparatus is being transferred from one molecule to another, thus spending a small fraction of its life in the free state. In the author's opinion, this view is scarcely tenable in this form, at any rate for low fields and high pressures. That positive charges are not transferred from molecule to molecule is shown from the experiments with radioactive recoil ions which retain their identity throughout the measurement.

§ 3. EFFECT OF VAPOURS

The effect of vapours with a marked dipole in lowering the mobility of the negative ion is a direct result of the theory. The permanent dipole increases the value of the clustering factor and the polyatomic molecules are of such dimensions as lead to an increase also in σ. This has been shown in some completeness by the



Effective size of ions formed of chain polar molecules.

author and L. R. Phillips* for the series of aliphatic alcohols which consist of long chain molecules of the form $CH_3(CH_2)_n.OH$, with a dipole roughly constant for the series and presumably unsymmetrically placed at the OH end. Table 1 illustrates the effect of adding alcohol vapour at 2 mm. pressure to dry air at atmospheric pressure. The table also shows the corresponding values for positive ions. The effect in the case of positive ions is much less and leads to the result that with the higher alcohols (and even for the lower alcohols at greater concentrations) the mobility of the negatives is reduced to a value lower than that of the positives.

Pictorially we might represent the effect for negatives as in the figure, which indicates how the molecules of a long chain polar compound may increase the effective size of the ion. In this case the permanent dipole and the induced dipole are additive. In the case of a positive ion, on the other hand, the sign of the attracting field would tend to reverse the orientation of the molecule, with the result that the binding force would be weaker; fewer alcohol molecules would be found in the layer at any given instant, and the net effect on the mobility would (in agreement with experiment) be considerably less.

Table 1

Added vapour (pressure = 2 mm.)	Mobility	
	Negative	Positive
None	2·15	1·38
Methyl alcohol ...	1·78	1·28
Ethyl ,, ...	1·46	1·26
Propyl ,, ...	1·16	1·23
Butyl ,, ...	0·88	1·05
Amyl ,, ...	0·86	0·96
Decane	1·98	1·38
Chloroform ...	1·26	1·34
Carbon-tetrachloride	2·0 (2)	1·34

Comparative results which confirm this view were also obtained with chloroform (polar) and carbon-tetrachloride (non-polar), and with amyl alcohol (polar) and the much longer molecular chain decane, $CH_3.(CH_2)_9.CH_3$ (non-polar), the non-polar substances having relatively little effect on the mobility of either sign.

§ 4. PURE GASES

The striking influence of polar impurities cited above is in itself sufficient to emphasise the importance of experiments in highly purified gases. It has been known for some time that, in certain instances where purity obtains, many negative ions do not take on a cluster but remain in the electronic state. This was first found to be the case at atmospheric pressure by Franck† for pure nitrogen, helium and argon, and by Chattock and Tyndall‡ for pure hydrogen. The addition of small traces of oxygen, however, destroyed the electronic character of the ion. Townsend and a number of co-workers have brought forward strong evidence that at lower

* Tyndall and Phillips, *Proc. R. S.* A, **111**, 577 (1926).

† Franck, *Verh. d. Dtsch. Phys. Ges.* **12**, 613 (1910).

‡ *Phil. Mag.* **19**, 449 (1910); *Nature*, **84**, 531 (1910).

pressures, of the order of several millimetres, the negative ion is always an electron in any non-polar gas, provided that all traces of polar impurity such as water or vapours from ebonite or tap grease be excluded; they regard the loading up of the electron in the experiments at atmospheric pressure as due not to oxygen but to water vapour.

This is a very interesting result, because it suggests that the clustering factor due to the induced dipole is a negligible quantity, and that, therefore, one should expect that vapourless oxygen or air should behave like nitrogen and give electrons even at atmospheric pressure. This has not yet been observed. Starr, Powell and the author* have looked for electrons in air at atmospheric pressure, taking the same precautions to avoid traces of vapour and with the additional advantage of an apparatus so designed that it was capable of separating the ions, if they were present, into groups of characteristic mobility. While a large number of electrons were found in nitrogen under these conditions, no success was obtained in air.

There is, however, a possible way of explaining the discrepancy, namely, that the air was sealed in a vessel containing polonium. Small traces of ozone and oxides of nitrogen are thereby produced which may act as a polar impurity. The concentration of such impurities will increase with the pressure; at low pressures it may well be negligible even with a continuous source of ionisation.

The results obtained at the Oxford laboratory at low pressure seem to the author to be significant in another respect. If in the absence of polar impurities the clustering factor is negligible because it is only an induced dipole which is operative, then the positive ion should under these conditions be monomolecular. No evidence of this has yet been obtained, but it is conceivable that owing to the longer life of the ion in the medium, even greater precautions to exclude impurity will be necessary before such evidence is forthcoming.

§ 5. POSITIVE IONS OF SHORT AGE

In the above we have been concerned with ions of relatively long age. There remains the problem of the nature of an ion at short ages. This has arisen from the investigations of Erikson by the method in which the ions are dragged by an electric field across an air blast. He showed that under these conditions there are two types of positive ion—initial and final ions as he termed them—and that the number of final ions grow with age at the expense of the initial ions. The mobility of the initial ion was found to be the same as that of a negative ion which, however, was not subject to the same transformation with age. The author and G. C. Grindley showed that the rate of transformation of initial ions into final ions is retarded by water vapour, the retardation increasing with increase of the pressure of the vapour, and they remarked† that observations in very dry air would be very desirable as the rate of transformation under these conditions might be so great as to be virtually instantaneous.

In order to confirm this view the experiment was repeated by the "four gauze"

* Tyndall, Starr and Powell, *Proc. R. S.* A, **121**, 172 (1928).
† *Proc. R. S.* A, **110**, 364 (1926).

method of Starr, Powell and the author, which as a closed vessel method permits of more thorough drying of the air. No initial ions were then found either in dry air or in pure nitrogen. Further confirmation has recently been obtained by Mahoney who used the air blast method with a closed circuit of dry air; he found no initial ions at an age of 0·0014 sec.*

This result taken alone would suggest that the initial ion is not a true ion of the gas under investigation, but one which arises from the presence of an impurity. The problem is, however, complicated by the fact that all impurities do not act similarly. Thus the author, Grindley and Sheppard have shown that the rate of transformation is greatly *increased* by ozone. The negative ion is unaffected under the same conditions.

The various possibilities may be put in the schematic form shown in Table 2. In this scheme the ions are called fast and slow instead of initial and final, and it is assumed that in all highly pure gases the negative ion is an electron.

Table 2

	Negative ions	Positive ions			
		Case A	Case B	Case C	Case D
Fast	ϵ	I_+	G_+	g_+	$g_+ + I_1$
Slow	I_-	G_+	$G_+ + I$	$g_+ + g$	$g_+ + I_1 + I_2$

ϵ is an electron.
I, a cluster of polar impurity.
G, a cluster of molecules.
g, a single molecule.

Case A. On this view the transformation would be due to the ion's shedding its impurity members and taking on gas molecules. The arguments against this explanation are:

(i) In view of the polar properties of the impurity it seems improbable.

(ii) One would expect that if it were correct the final ion would be the faster of the two; the reverse is the case.

(iii) All impurities do not act similarly. (See above.)

(iv) It does not explain the equality of the slow negative and fast positive mobilities, because the two clusters would probably differ in size.

Case B. In this case the transformation is due to the collection of impurity by occasional collisions. Indeed, the relatively slow rate of transformation suggests some special type of collision of relatively rare occurrence. In this case the equality of mobilities of slow negative and fast positive ions would be a coincidence. But the action of a polar impurity water vapour in favouring the persistence of fast positive ions does not fit with this view.

Case C. This is the suggestion put forward by Erikson himself. It requires, however, the abandonment of our views of a cluster, and at present it has no theoretical support.

* *Phys. Rev.* **33**, 217 (1929). Mahoney remarks in his paper that this result is contrary to the work of the author and Grindley. Reference to the original paper will show that this statement is incorrect.

Case D. This alternative is based on the assumption that the true positive ion of the gas at atmospheric pressure is monomolecular and that no samples of gas have yet been experimented upon which are free from some very active impurity I_1. The transformation would then be due either to the substitution or to the addition of a second impurity I_2. There is no evidence at present for this view, but it is possible that a still greater degree of purity is essential for the test.

§ 6. FUTURE PROGRESS

Such difficulties as these convince the writer that the most fruitful line of attack at the present time is so to improve the technique that the nature of the ions, both positive and negative, in pure monatomic and diatomic gases at atmospheric pressure may be investigated under the best conditions of modern vacuum and gas practice. Some surprise may possibly be felt that this has not already been done. But it must be remembered that elimination of small traces of impurities and other experimental precautions become more difficult as the pressure is raised.

With this end in view a new apparatus has been constructed by Mr C. F. Powell in the Wills laboratory at Bristol. Preliminary results of a suggestive character have already been obtained and we are hopeful that they will clear up a number of discrepancies and bring the work into closer harmony with that which has already been carried out in pure gases at low pressures. Not until this is completed can any theoretical examination be conducted upon a secure foundation.

When this information has been obtained, it may still be necessary, in order to discover the real nature of an ion at high pressures, to make experiments by new methods which aim at measuring the mass of an ion and not its mobility.

An attempt in this direction was made in carefully purified hydrogen as far back as 1908 by Chattock and the author*. The method depended upon the measurement of the minute reduction in pressure set up by the absorption of the ions by a platinum electrode. The results suggested that the positive carrier, which was absorbed, was a singly charged molecule. On the cluster theory this would mean that the remaining neutral molecules were, so to speak, brushed off at the surface and so rendered free to evaporate. The experiment, however, must now be reviewed in the light of our present-day knowledge of surface gaseous films, and the conclusions cannot be regarded as proven.

It is conceivable that the methods applied by Smyth to positive rays at low pressure might be applicable to this problem, but the experimental difficulties would be greatly complicated by the higher pressures required, and the necessity of avoiding such collisions with the molecules of the gas as would lead to the disintegration of a cluster before the ion reached the mass spectrograph.

NOTE BY C. F. POWELL

The method employed in the new apparatus, referred to by Prof. Tyndall as having been constructed by me, is the "four gauze" method. This has already been described by Tyndall, Starr and myself†, and was described independently in a

* *Phil. Mag.* [6], 16, 24 (1908). † *Loc. cit.*

modified form by Mr van de Graaff*. It may be regarded in a sense as analogous to Fizeau's method for determining the velocity of light. In the form in which we are using it, it has high resolving power in that, if more than one type of ion is present, each type is represented by a sharp peak in a current-frequency curve, the height of the peak being proportional to the number of ions of that type present in the gas. The age of the ions under investigation is also controllable over a considerable range.

In its earlier form the apparatus was all of metal and was enclosed under a bell jar with external wax seals. In a test with nitrogen it was found that while a large number of the negative ions were electrons, some ions of mobility given by the cluster theory were also present. In its new form the apparatus is free from wax seals and is enclosed in a glass tube. The leads from the various electrodes pass through glass pinches, and the one communicating with the electrometer through a graded pyrex seal. The whole can be placed in an oven and baked out before admission of the pure gas.

Under these conditions of greater purity, we anticipate that we shall obtain nothing but electrons as the negative ions. If this be the case, we are also hopeful that the investigation of the positive ions may be of some interest because, as Prof. Tyndall states, theory suggests that the positive ion may be monomolecular in the absence of all polar impurities. The plant for similar investigations in pure helium is also now available.

DISCUSSION

Mr J. H. Awbery. I was most interested in Prof. Tyndall's very clear summary of some of the points which present difficulties in this subject and, like others, I shall look forward with eagerness to the results of the work he is now undertaking on mobilities in very pure gas. There is one objection which has sometimes been urged against any cluster theory of ions, on which some of those present may be able to throw some light. Although there is some disagreement as to the facts, I rather gather from the literature that mobilities do not increase when the electric field is made very high. On a cluster theory, one would expect that sufficiently violent impacts would break up the cluster and lead to a higher mobility.

I would suggest that light might be thrown on some of the problems by further study of the large ions of mobilities about 1/150 and 1/3000, which were discovered and studied by Langevin and Pollock, and which seem of late to have received little or no attention.

Mr S. P. MacCallum. I should like to congratulate Prof. Tyndall on the great advance in technique he is bringing to the measurement of the mobility of ions at high pressures. His remarks have confirmed the opinion I have had that very few reliable conclusions can be deduced from most of the earlier work. Not much progress can be made until the experiments are made in pure gases and in apparatus in which the gases can be kept reasonably pure. The minutest traces of impurity can in some cases produce a very large effect upon the results.

* Van de Graaff, *Phil. Mag.* **6**, 210 (1928).

We have found at the Electrical Laboratory, Oxford, that a high frequency discharge in monatomic gases is a very effective means of removing impurity when the amount is small. In connection with this method we have also found that amounts of impurity which we were unable to detect with a spectroscope can be detected easily by a measurement of the sparking potential. Circulation over charcoal at liquid air temperature is a lengthy process at high pressures and Prof. Tyndall may find the method suggested very useful.

Mr L. G. H. Huxley. Mr Awbery has suggested that perhaps under the action of a large electric force the cluster ion would tend to break up, and I believe that some experiments that I have carried out on the corona discharge in helium and neon give some information on this point.

By application of a formula developed some years ago by Townsend it is possible to make an accurate measurement of the mobilities of ions for large values of the ratio of the electric force X to the pressure p occurring in the corona discharge. My experiments showed that for values of X/p of the order of ·5 to 2 the mobility of the positive helium ion, when reduced to atmospheric pressure, was about 30 cm. per sec. This is the order of magnitude given by Langevin's formula for a monomolecular ion and is much greater than the number 5 given by Franck and Pohl for small values of X/p. Similar results were obtained in neon and nitrogen. The positive ions in neon and helium also exhibited an increase in mean free path with increasing energy.

Dr D. Owen. It is a pleasure to hear further of the interesting work on ionisation that is being carried out by Prof. Tyndall and his pupils at Bristol. It is interesting to note that the work on mobilities at normal pressures leads to the definite conclusion that the ions in gases are not single molecules but aggregates. It would be of interest to know whether in the case of positive ions a transition occurs at some stage in reduction of pressure, of the kind which is known to occur in the case of negative ions, being here interpreted as indicating that the ions are electrons which maintain their existence in the free state. Again, from the structure of the polar molecules of the alcohol series CH_3OH, etc., with their exterior positively charged atoms, it might be inferred that the negative ions in a gas containing vapours of these alcohols should be larger and hence of lower mobility than the positive ions, which is contrary to the usual order. Is this actually the case? The increasing importance at the present time of power transmission at very high voltages prompts another question: whether Prof. Tyndall can give any data as to the character of the ions which occur in high-tension discharges?

Dr F. J. W. Whipple. A remark made by Mr Awbery seemed to imply that no work had been done in recent years on the large ions discovered by Langevin. That is far from being the case. Prof. Nolan and his school in Dublin have devoted much time to the investigation of the ions of different kinds in the atmosphere. They have demonstrated that the Langevin ions are Aitken nuclei. Roughly speaking, one-third of the nuclei present at any time are neutral; the other two-thirds carry charges positive or negative. These charges are acquired from the

small ions. When the atmosphere is polluted, nuclei are numerous and the small ions are short-lived; conductivity is therefore low and potential gradient assumes high values. The interplay between large and small ions is one of the most important phenomena in atmospheric electricity.

Reply by Professor TYNDALL: In reply to Mr Awbery, I think that the effect of polar impurities may be important at high field strengths, because, owing to the strong attraction, their molecules would be less readily removed from the cluster. The experiments in pure gases cited by Mr Huxley are very significant in this respect. Loeb's deductions from the early form of Langevin's formula and an expression due to J. J. Thomson are based upon a misinterpretation of Thomson's paper. Thomson's paper is an attempt at a statistical treatment of the subject based upon mean free paths; it has not yet, however, been developed to a numerical stage. In the course of the paper he deduces two expressions for the reciprocal of the energy free path, one assuming only elastic collisions, and the other assuming only attractive forces between ions and neutral molecules. These terms must, of course, be added, but Loeb omits the former from his calculation and thereby obtains an expression independent of σ.

There is a corresponding special case of Langevin's 1905 formula which is derived from a treatment of elastic collisions coupled with attractive forces but without involving mean path considerations. Langevin's expression may be written

$$\text{Mobility } (k) = (1 + m/M)^{\frac{1}{2}} . A/\sqrt{\rho (K-1)},$$

where A is a function of a parameter, the value of which depends upon the relative importance of elastic collisions and attractive forces. When the latter are of much greater importance than the former, A tends to a limiting value 0·505.

With reference to Mr MacCallum's suggestion of the use of high frequency discharge for cleaning up impurities, I might say that we had considered this possibility, but doubted whether it could be applied to gases at the high pressures we are using. His opinion, however, that it may be practicable even at 100 mm. is very welcome, and brings the method within the range of practicability.

In the case of high tension discharges from transmission cable raised by Dr Owen, we have, presumably, in the glow itself, ions of very short age travelling in strong fields, with, however, complications due to the presence of a high concentration of polar impurities formed by the chemical action of corona discharges. Dr Owen is correct in his surmise that, in the presence of organic vapours, the positive ion may move faster than the negative. I did not state this at the meeting, but reference is made to it in the printed text of the address.

The matter of large ions to which Mr Whipple referred was one of the various subjects omitted for the sake of brevity from the paper. It is an interesting subject of meteorological importance, and bound up with the question of condensation nuclei in the atmosphere. It is full of experimental difficulties, the chief of which is that of obtaining a controllable supply of ions of the same kind which can be repeated indefinitely in a series of experiments.

THE TEACHING OF GEOMETRICAL OPTICS

Discussion held on April 26, 1929

I. THE NEED FOR REFORM IN THE TEACHING OF OPTICS

By T. SMITH, M.A.

ABSTRACT. The student should not be taught any theory until he has learnt to distinguish between light and sight, and acquired a knowledge of the properties of light, and of the effects observable with optical instruments. In his practical work he should realise the limited powers of unaided and of aided vision, understand why there is always a limit to resolving power, and why numerical aperture is important. He will then realise that the geometrical theory cannot do more than locate "perfect" images. Examples of the failure of this theory outside these limits are given. The wave treatment should be taught as a reliable way of dealing with optical problems generally. The use of simple matrices is advised. In advanced work Hamilton's method is preferred. The sign conventions are as in ophthalmic practice and in analytical geometry.

THERE should be no need to-day to urge the view that no scheme of education which ignores optics can be deemed satisfactory. If reasons for this belief are demanded we have only to mention the importance to each of us in our daily lives of making a right use of the basic facts of vision and of the elementary properties of light; our dependence for many purposes on optical instruments, such as spectacles, cinematograph projectors, field glasses, microscopes; and not least, for those who desire to understand the world around them, the fact that in optics are found some of the more intelligible applications of those ideas, still somewhat strange to those who were learning physics in the 19th century, which seem destined to affect so profoundly the views of that world revealed in the pictures now beginning to be unfolded before us, to make the justice of our claim apparent. The list need hardly be extended. We ought however to observe that the chances of anyone reaching, without instruction, correct views on any one of these items are negligibly small. The need for instruction exists, and since it is not merely to a few specialists that the matter is of interest, much of this instruction should be given in the schools. The first question then to be discussed is the form this elementary instruction should take.

When our Hon. Secretary first mentioned the need for a change in existing teaching, and invited me to give the opening address at this discussion, I naturally asked what optics was taught in schools to-day. His reply was brief but illuminating: "$\frac{1}{u}+\frac{1}{v}=\frac{1}{f}$." I shall have something to say on this equation when I come to deal with lenses. At the moment I only wish to point out that it represents that aspect of lens theory which is least interesting and least instructive. In this discussion we are dealing with the wider subject of optics, and it is not unimportant to observe that optics does not lie entirely within the domain of physics. Physiological optics and the psychology of vision are more interesting and more important

to most people than laws about optical images, and, at all events in the initial stages, these, together with the outstanding physical properties of light, should form the foundation of the course, which ought to be experimental. Perhaps the most important aim at this stage should be to teach the pupil to think correctly about light and sight; the experiments involve both physical stimulus and mental sensation. Our aim is to learn to reason accurately about the former through our experience of the latter. The laws of these two realms differ, and the task is not as simple as the student may at first be disposed to think. The experiments will generally be qualitative rather than quantitative. Experiments on brightness and colour, persistence of vision, fatigue, glare, and the power and limitations of adaptation to changes in external conditions, with their relation to the exciting light, are typical of the work that might be done early in the course.

In this preliminary stage the student should have realised that the eye is very fallible as a measuring instrument. When it must be employed in this capacity care is taken to use it under the conditions which have been found to suit it best, and ingenious instrumental aids are nearly always employed. The evolution of photometers, to take only a single example, affords a good example of the way in which we have learnt to make effective use of a somewhat indifferent instrument. We have here reached the stage when light has to be considered as a form of energy. The discussion of the two forms in which energy can be radiated, corpuscles and waves, introduces a part of the course which can be well illustrated by experiment. The apparatus used is perhaps not often found in schools, but it need not be expensive. For example replicas of diffraction gratings, suitable for illustrating the periodic character of light and the relation between colour and wave length, can be obtained for a shilling or two.

The student will usually have sufficient qualitative knowledge of the properties of optical instruments like the telescope, microscope, and photographic lens for these to be employed in experiments set up by the master. The aim in the next stage, and the one to which I attach as much importance as any, is to teach the student how to use these instruments correctly. There is here a difficulty which can only be met in the course of time. The proper use of the microscope, particularly with objectives of high power, is of the utmost importance for the advancement of knowledge in several directions, and ought to be taught in school; but there are no teachers. The trouble is that most microscope users are not aware that they do not use their instruments properly—i.e. so as to see everything that the system is capable of showing and avoid false effects (the interpretation of what is seen need not detain us here). Indeed it has been said that all the expert microscopists in the country can be counted on one's fingers. It is chiefly a matter of using the instrument intelligently, taking trouble, exercising patience, and gaining experience. The teacher should have a good practical knowledge of the wave theory of light, and this points to the physics master as the best instructor. A holiday course for science teachers on the use of the microscope, by one of the great masters of the art, would probably be the simplest means of bringing about an improvement. However, we are counting our chickens before the eggs are laid. The microscope and telescope

are used to enable us to see what is beyond the power of our unaided eyes. The first step in a course on the use of instruments therefore consists in experiments on the power of the eye in direct unaided vision. The most important facts to be brought out concern the capacity of the eye to distinguish detail under different conditions. The need for sufficient illumination may be exhibited in many ways. Next the importance of the size of the aperture admitting light to the eye should be considered. If this is reduced much below the natural size of the iris, detail disappears even though the illumination is ample. Quite striking observations may be made with the simplest apparatus. For instance objects containing prominent straight lines lying in various directions may be viewed through a straight cut made in the film of a blackened dry plate. Lines parallel to the cut will appear blurred or even disappear entirely, while those approximately perpendicular to the cut can be seen distinctly. This experiment, with suitable variations, should be used to show that the discernment of detail is dependent upon the use of a light-restricting aperture of considerable size, and that as the aperture is reduced in any given direction the size of the smallest detail we can distinguish in that direction changes in the inverse ratio. A familiar example is afforded by the ability of a photographic lens, which has an aperture much larger than that of the iris of the eye, to record on a photographic plate detail which is invisible to an eye placed in the same position. We also infer, contrary to the generally received opinion, that when a photographic lens is stopped down, the images of objects in sharp focus lose, instead of gaining, in detail. These conclusions will prepare the student to understand the importance of numerical aperture in deciding the smallest detail that can be correctly rendered in an image. The ability of the eye to detect departures from exact alignment of two parallel straight lines, where they cross a third line, should also form the subject of experiment, as the difference between this magnitude and the separation of two points discernibly distinct is of considerable significance.

When instruments are introduced to extend the powers of observation of the naked eye, the experiments on the effect of reducing the size of the aperture should be repeated to demonstrate that the same laws hold. It is particularly important for the student to realise the uselessness of increasing the magnification beyond the value suitable for the aperture actually effective in the experiment. In all these experiments the size of the exit pupil should be measured. The character of the images of simple objects may next be investigated by examination under a low-power microscope. The series should include the image of a star seen in a good telescope, with the objective used both at full aperture and stopped down; and the appearance of double stars on both sides of the resolution limit: in the microscope parallel rulings of various spacings should be examined, and the student should be led on to appreciate the importance of correct illumination. Incidentally the microscope, or rather the microscopes, for there should be several, may be used to exhibit natural history or mineral specimens of optical interest, illustrating for example brilliant coloration due to the optical effects of particular kinds of structure.

I have suggested that the student should gain, in some such way as that just

sketched, a knowledge of the principles involved in the use of optical instruments as well as practice in the art itself, before he is confronted with the special theory of optical instruments. If we have to choose between teaching the use of an instrument and teaching a theory which rarely calls for application, by all means let us select the former. Far more of our students will have occasion to use the microscope as a tool than will ever need to calculate conjugate distances for a thin lens. Knowledge of the kind that has been mentioned is essential for the proper handling of an instrument, but no one would be any the poorer if the equation said to represent present teaching were entirely forgotten and lost. It should be unnecessary to add that the lenses of the telescope and microscopes used in the experiments should be of really good quality. With the microscope particularly it is possible for poor equipment to be worse than none.

Some of the most important properties of light considered as a wave motion have yet to be introduced before we can consider the theory of lenses. They include the constancy of the speed with which light travels, the principle of interference, the qualitative explanation of rectilinear propagation based upon interference, and the dependence of the speed upon the medium and the frequency (which is invariable). From these we can construct a complete explanation of reflection and refraction. The very great value of the speed should be pointed out on account of its importance in leading us to regard our sensations as a function of the stimuli in a statistical sense only. I assume that the instruction now given on all these points is satisfactory.

It may be thought that with the deduction of the rectilinear propagation of light from wave concepts, and the derivation of the laws of reflection and refraction in their geometrical form, the way becomes clear for the discussion of the properties of lenses from the geometry of rays, in the manner with which we are all familiar. I wish to urge the contrary view. I am not aware that anything of physical importance is gained, though I am conscious that much is lost, by introducing the ray idea into the subject. We are in effect adopting, for use in two ill-defined and overlapping sections of optics, two mutually inconsistent theories of light. There might be something to be said for the retention of the geometrical theory in elementary teaching if it possessed greater simplicity, or if it presented a picture more readily grasped by the student's mind: in fact the advantage lies on the other side. In technical optics ray-tracing is an important process, but this does not justify the use still made of rays in teaching the principles of image formation. The ray theory leads to a number of conclusions which are contradicted by experiment. On the other hand observation confirms the predictions derived from the wave theory. If, notwithstanding these facts, we continue to accept the geometrical doctrine, we are false to the professed principles of experimental science. As an exercise in geometry the ray theory is not out of place: as a branch of physics, or in association with "optical," we are bound to object to its use.

The errors which inevitably accompany the geometrical theory are due to the assumption that the direction in which the radiant energy falling on a vanishingly small area of an arbitrary surface is propagated, depends only on the direction of

the normal to the wave surface where it meets this area. The explanation of rectilinear propagation according to the wave theory requires comparatively remote parts of the wave front to exercise some influence, and in fact the explanation breaks down near the edge of a wave or when the wave front is not extensive. This agrees with the observation that detail is lost when the aperture limiting the light traversing the instrument is reduced in size. The geometrical laws of refraction and rectilinear propagation may therefore mislead us through confining our attention to too restricted a part of the wave front. We can readily illustrate the kind of error that a student would be apt to make. Suppose we have a solid circular cylinder, with flat ends normal to the cylinder axis, of a transparent material in which the refractive index is a regular function of the distance from the axis. Consider a ray which enters the cylinder in a direction parallel to the axis. It is undeviated on entering the cylinder because the incidence is normal. The refractive index in the direction in which it is travelling is constant, and the geometrical theory would lead to the conclusion that it would continue to travel through the cylinder in its original direction. The wave theory leads to a very different conclusion. As the refractive index differs at different distances from the axis, the speed of the wave front depends on this distance, that is to say the wave front becomes progressively more and more inclined to its original direction, or the energy is propagated in a curved path. If the refractive index is a maximum along the axis and falls continuously as the radius increases, the path will oscillate from one side of the axis to the other. This example incidentally illustrates the ease with which, on the wave theory, we can discuss the effects of non-homogeneity in our materials—a problem which arises frequently in practical work. As a second illustration consider a wave front after refraction by a lens in the form of a portion of a perfect sphere contracting as it advances towards its centre. In general the area of the wave front when refraction occurs will be sufficiently large to ensure effective obedience to the laws of refraction and rectilinear propagation. As the wave advances the dimensions of the front decrease, and before the focus is reached it has shrunk so much that the conditions necessary for rectilinear propagation are no longer satisfied. We see therefore that the very fact that the light is converging towards a focus implies the failure of those conditions upon which the ray theory relies for its validity. In other words geometrical optics is incompetent to deal with the very phenomena for which alone it exists. It misleads us if we consider by its aid the nature of an image; it indicates the converse of the true effect when the size of the aperture is varied; it gives incorrect laws for the depth of field and the depth of focus; it assigns the wrong position for the best image in the presence of slight aberration; in short it is much more difficult to find a problem to which it yields the correct answer than to extend the list of those in which it contradicts the wave theory. I suggest, in view of these notable failures, that it is essential for us to abandon the geometrical theory of image formation.

There is one method of treating the lens problem, without utilising any theory of light, on which a word of criticism may be made here. The argument used is that every point (let us say on the axis) of the object space has a unique image in the

image space, and *vice versa*. The relation between real co-ordinates is thus one-one, and is therefore of the form

$$axx' + bx + cx' + d = 0,$$

where a, b, c, d are constants. This argument is not valid, for we still obtain a one-one relation between the real co-ordinates if we replace x and x' by x^m and x'^n where m and n are any odd integers.

We have now to consider how to present the theory in a manner free from the objections that have been raised against the ray treatment. We shall clearly avoid these difficulties if we consider the wave as a whole and base the theory on the fundamental properties of a wave motion. In some respects the exact theory is simpler than the approximate relations used in elementary discussions; as the exact theory requires the use of mathematics somewhat too advanced for some students, we will only consider the elementary treatment here.

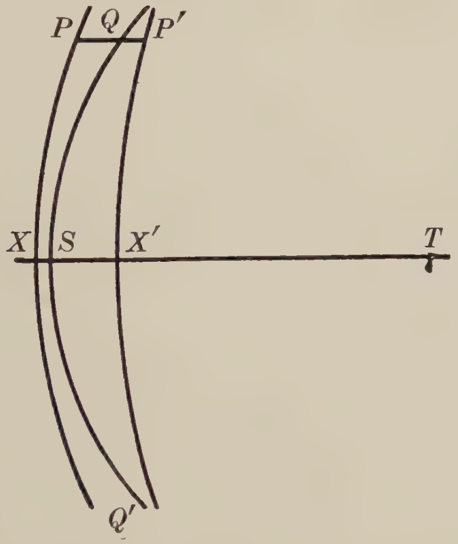


Fig. 1.

Suppose that QSQ' (Fig. 1) is a section of a refracting surface which forms part of a sphere with its centre at T. The curvature R of this surface is measured by the reciprocal of the radius ST : R is to be considered positive when the incident light passes through the surface from the convex to the concave side. Let PX represent in section a spherical incident wave, having its centre of curvature on the straight line XST, just before refraction; and similarly let $P'X'$ represent the wave in section just after refraction. From the symmetry of the incident wave and the refracting surface about ST we can say that, if the media on either side of QSQ' are homogeneous, the emergent wave will also be symmetrical about the same axis. Thus the equations to the three surfaces may be written in the form:

Incident wave	$x = x_0 + \frac{1}{2}y^2U +$ terms in $y^4, y^6, \ldots\ldots,$
Refracting surface	$\xi = \xi_0 + \frac{1}{2}y^2R +$ terms in $y^4, y^6, \ldots\ldots,$
Emergent wave	$x' = x_0' + \frac{1}{2}y^2V +$ terms in $y^4, y^6, \ldots\ldots,$

where U and V are the curvatures (bearing the signs of similarly oriented refracting surfaces) near the axis of the incident and emergent waves. Now the light travels along the axis from X to X' through S: let us suppose that PQ_1P_1' is the stationary path from P to the emergent wave. By the properties of a stationary path any near path will differ in length from the stationary path by a negligibly small quantity. We therefore make a negligible error by substituting PQP', a straight line parallel to the axis, for the actual stationary path. As the time taken to go from P to P' is equal to that required by the light to travel from X to X', we have, since the time needed to traverse a given length is proportional to the refractive index,

$$\mu . PQ + \mu' . QP' = \mu . XS + \mu' . SX',$$

or

$$\mu (\xi - x) + \mu' (x' - \xi') = \mu (\xi_0 - x_0) + \mu' (x_0' - \xi_0),$$

or, since the y's for all the surfaces are equal,

$$\tfrac{1}{2} y^2 \{\mu (R - U) + \mu' (V - R) + \text{terms in } y^2, y^4, \ldots\} = 0.$$

We *assume* that we commit no serious error by supposing all the surfaces to be strictly spherical. We can then neglect the terms involving y in the bracket, since the correct result is determined by the leading terms. Moreover the range for y is finite, and we thus reach the equation of refraction

$$\mu' V - \mu U = (\mu' - \mu) R = \text{constant} \qquad \ldots\ldots(1)^*.$$

This constant is called the power of the surface. Equation (1) represents a change in the curvature of the wave which occurs as the light passes through the refracting surface. The effect is easily visualised.

If now the emergent wave encounters a second refracting surface of curvature S separating media of indices μ' and μ'' immediately after it has passed through the first surface, the curvature of the wave on incidence at the second surface will be sensibly the same as that with which it emerged from the first surface, and the equation for the second refraction will be

$$\mu'' W - \mu' V = (\mu'' - \mu') S = \text{constant}.$$

Adding this to the first equation we obtain

$$\mu'' W - \mu U = (\mu' - \mu) R + (\mu'' - \mu') S = \text{constant},$$

which is essentially of the same form as (1). It follows that (1) may be at once generalised to apply to any number of refractions at surfaces crossing the axis in the immediate neighbourhood of a single axial point. Powers are merely additive. In the important case when the external media are air, for which the refractive index is assigned the value unity, the equation becomes

$$V - U = F \qquad \ldots\ldots(2).$$

It may be observed that when light travels in the reverse direction the signs of the curvatures R, U, V are changed, and also the various media are encountered in the reverse order, causing an additional change of sign. It follows that the equations are independent of the direction in which the light travels.

* The systematic treatment of lens theory by this "curvature method" was initiated by Prof. S. P. Thompson, and extended by Dr C. V. Drysdale.

Equations (1) and (2) are specially simple forms of the linear equation

$$a + b.\mu U - c.\mu' V - d.\mu U.\mu' V = 0 \qquad \ldots\ldots(3).$$

This equation remains linear if we convert it from a relation between curvatures into a relation between lengths, viz.

$$a.\frac{u}{\mu}.\frac{v}{\mu'} + b.\frac{v}{\mu'} - c.\frac{u}{\mu} - d = 0 \qquad \ldots\ldots(4).$$

The u and v in this equation are the distances of the centres of curvature of the incident and emergent waves from the vertex S of the refracting surface. Since we suppose that the waves are spherical, they will remain spherical, although of changing curvature, as they travel onwards. Thus when the incident wave was still a distance t from the vertex of the first refracting surface, the front was a portion of a sphere of radius $u_1 = u + t$, and similarly after travelling a distance t' from the vertex of the last surface the emergent wave will form a portion of a sphere of radius $v_1 = v - t'$. If we substitute from these equations for u and v in (4), we obtain an equation which we may write

$$a'.\frac{u_1}{\mu}.\frac{v_1}{\mu'} + b'.\frac{v_1}{\mu'} - c'.\frac{u_1}{\mu} - d' = 0 \qquad \ldots\ldots(5),$$

where a', b', c', d' are functions of a, b, c, d, t/μ, t'/μ'. We may now reconvert (5) into an equation connecting the curvatures U_1 and V_1 of the incident and emergent waves at distances t and t' from the first and last surfaces respectively, viz.

$$a' + b'.\mu U_1 - c'.\mu' V_1 - d'.\mu U_1.\mu' V_1 = 0 \qquad \ldots\ldots(6).$$

In this form the equation is suitable for taking into account refractions where the wave curvatures are U_1 and V_1 respectively, leading to new constants a'', b'', c'', d'' which depend only on a', b', c', d' and the powers of the new refractions. Obviously the process may be indefinitely extended, and equations corresponding to (5) and (6) may be found for any system of refracting surfaces symmetrical about an axis. Now a'', b'', c'', d'' are constants, that is to say they are independent of the curvatures the incident and emergent waves happen to possess. It follows that the properties of the system of lenses are contained in the values of these constants, and to find these properties we only need to be able to calculate the constants. One method, but a very poor one, of obtaining them consists in finding step by step the values of the final V's for arbitrarily selected initial U's. I suggest that it is a mistake to lose so favourable an opportunity of introducing the pupil to simple matrices. They represent by far the least laborious method of finding the properties of the system, and their use should excite the interest of the student in a manner which is not otherwise easy when the algebra is of a rather monotonous kind. It is readily seen that when the systems a_1, b_1, c_1, d_1 and a_2, b_2, c_2, d_2 are combined to form the compound system a_{12}, b_{12}, c_{12}, d_{12}, the relations between these constants are

$$\begin{pmatrix} b_1 & d_1 \\ a_1 & c_1 \end{pmatrix} \begin{pmatrix} b_2 & d_2 \\ a_2 & c_2 \end{pmatrix} = \begin{pmatrix} b_{12} & d_{12} \\ a_{12} & c_{12} \end{pmatrix} \qquad \ldots\ldots(7).$$

Of the four components of any matrix, b and c are pure numbers, a is a power, and

d is a length—Cotes' "apparent distance" with its sign changed. Thus for a thin lens or a single surface the matrix is

$$\begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix},$$

and for a displacement of the point where the wave curvature is measured through the non-virtual distance t the matrix is

$$\begin{pmatrix} 1 & -t/\mu \\ 0 & 1 \end{pmatrix}.$$

Thus the four constants of a single lens of refractive index μ and axial thickness t, with surfaces of curvatures R and S respectively, are given by the product

$$\begin{pmatrix} 1 & 0 \\ R(\mu-1) & 1 \end{pmatrix} \begin{pmatrix} 1 & -t/\mu \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ S(1-\mu) & 1 \end{pmatrix},$$

the lens being surrounded by air. By inspection of the elementary factors of the general matrix we see that

$$\begin{vmatrix} b & d \\ a & c \end{vmatrix} = 1.$$

We may modify the generalised equation for the curvatures of the wave surfaces by writing X and X' for U and V respectively. The preferred form for the equation is

$$(\mu X \quad 1) \begin{pmatrix} b & d \\ a & c \end{pmatrix} \begin{pmatrix} 1 \\ -\mu' X' \end{pmatrix} = 0,$$

and this is always valid unless X and X' are large, corresponding to considerable shrinkage of the wave front. Now according to the wave theory there is no such thing as a point image, so that there is some objection to transforming this into an expression for the distances x and x' of object and image from the external reference points for which the central matrix is constructed. So long however as we exclude all but spherical waves, symmetry entitles us to claim the centre of curvature as a mean point about which the image will be formed. With this understanding we obtain the relation which is most useful in practical work, viz.

$$(1 \quad x/\mu) \begin{pmatrix} b & d \\ a & c \end{pmatrix} \begin{pmatrix} -x'/\mu' \\ 1 \end{pmatrix} = 0.$$

We have so far only considered a wave which meets each refracting surface normally at its vertex, but we can see at once that the equations are unaltered for waves making a small angle with the refracting surface at the vertex. From symmetry the relations satisfied by such waves on refraction are independent of the sign of the small angle, and the relation is thus stationary as the angle passes through the value zero. It follows that the difference between the relation for small angles and that which we have already considered is insignificant. The same conclusion may be reached by noting that each separate refracting surface is symmetrical about its centre.

To complete our theory we require equations to show the relative distances of an object and its image from the axis of symmetry. From the symmetry of a single spherical surface about its centre we readily derive the relation

$$\frac{y}{y'}=\frac{x-r}{x'-r}=\frac{X'(R-X)}{X(R-X')}=\frac{\mu' X'}{\mu X}.$$

It is easy to generalise this in the form

$$a=\mu X(-b+1/m)=\mu' X'(c-m),$$

where m is the transverse linear magnification y'/y.

The process we have followed of first forming equations from the physical principles of the subject, and subsequently separating the variable physical quantities—in this case the curvatures of the waves as they pass through specified points—from the constants of the system, and calculating the latter from the constructional data of the instrument, is of wide application. The student should realise the importance of following this plan wherever possible.

The familiar expressions for the positions of conjugate points relative to the cardinal points of the system are easily found if they are required. In practice nothing is gained by so doing. The positions of conjugate points with respect to the external surfaces of the system is the factor which determines the utility of the instrument. In laboratory measurements it should be obvious that greater accuracy is possible when measurements are made from the vertices of rigid surfaces than when they start from points, of which at least one is almost invariably inaccessible, that have to be located by means of their theoretical properties.

We may stop a moment to consider the expression of this class which occurs most commonly in elementary text-books. When the two reference points are conjugate to one another the apparent distance of one from the other vanishes, or $d=0$. The equation connecting the curvatures (assuming the system to be immersed in air)

$$a+bU-cV=0$$

is obviously preferable to that connecting distances, which may be written

$$a+\frac{b}{u}-\frac{c}{v}=0.$$

Whatever conjugate points are chosen we have $bc=1$, and if the unit points are selected the values are $b=c=1$. The two equations then become

$$V-U=a \qquad \text{......(8)}$$

and

$$\frac{1}{v}-\frac{1}{u}=a.$$

If we compare these with the equation mentioned at the beginning of this address, we may note that the physical significance of the equation is obscured by the use of distances instead of curvatures; that by adopting the focal length as the fundamental constant instead of the power useless arithmetical operations become necessary,

and the extremely simple law for combining lenses in contact is replaced by one which precludes reliable mental estimates of the properties of the joint system; and that, by measuring distances in opposite directions on the two sides of the lens, an unnecessary and undesirable departure has been made from the practice which mathematical experience has proved to be desirable. In consequence of this change of sign the meaning of the simpler equation, that a lens increases the curvature of a wave by a constant amount, would not be evident if curvatures were used instead of distances. It is much to be desired that this particular equation should be no longer used: it should invariably be replaced by equation (8), with the dioptrie (or dioptre) as the unit of curvature and of power. The signs for powers universally employed in ophthalmic optics—plus for a converging lens, minus for a diverging lens—should be adopted.

After the student has become acquainted with the laws of optical instruments he should be shown how to prove that the apparent brightness of an extended object cannot be increased by optical means. The fact itself can with advantage be mentioned earlier in the course. For example he should be told to notice that the moon looks no brighter when seen through a telescope than when viewed by the naked eye: on the other hand a telescope greatly extends the number of stars that are visible.

I have already indicated that there is little or nothing to be gained by retaining the so-called cardinal points as those to which all measurements must be referred. Associated with these points are corresponding planes—focal planes and unit planes—the chief function of which is to facilitate graphical constructions for the paths of rays refracted through the system. It goes almost without saying that if we have to abandon the ray concept in teaching optics we shall have little use for these surfaces. But that is not the only reason for thinking there would be a definite gain in rejecting them. The teaching now given is not content with saying that there are focal surfaces and unit surfaces; it goes on to assert that these surfaces are planes, i.e. that their curvature is zero. Now the slightest consideration shows that this is not true. When the optical system consists of a single surface it is obvious that both unit surfaces coincide with the refracting surface itself: that is to say their curvature is greater than the power of the system itself. Again if we take a plano-convex or a plano-concave lens, the unit surface on the one side coincides with the curved surface, and on the other it is the image of this curved surface in the plane. Both curvatures are large, and they are unequal. We may assert as a general proposition that the surfaces are decidedly curved. The doctrine of the planes has done much, I think, to conceal the interest which should be taken in optical systems, and in particular to prevent proper appreciation of the very notable achievements which have to be placed to the credit of lens designers in recent years. I must content myself here with pointing out that the focal surface for light refracted in the vicinity of a given point on the surface of a sphere is spherical, and that as we move the point along the surface the spherical image locus moves. It is at once evident that there is enormous aberration in the image when the aperture is large, and that the task of bringing all these loci into coincidence to secure good definition, and meeting

the additional condition that this common locus, contrary to all the inclinations of the constituent lenses, is to be plane and not spherical, is one that demands great knowledge, patience, and skill. If the real facts were more fully realised, we should be as ready to recognise in the finest high-power microscope objectives and in the best rapid wide-angle photographic lenses the consummate skill of the designer, as we are to appreciate in the largest telescope discs the triumphs of the glass-maker.

I have given some indication, I fear very sketchily and imperfectly, of the way in which I think we could advantageously change the elementary teaching of optics. If anyone doubts the advantages of using matrices in optics I would suggest for his consideration the problem of an assemblage of astigmatic surfaces with their powers, their orientations, and their separations arranged haphazard. The matrix solution is of the simple form

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \psi_1 & \pi_1 & 1 & 0 \\ \pi_1 & \phi_1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\tau_1 & 0 \\ 0 & 1 & 0 & -\tau_1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \psi_2 & \pi_2 & 1 & 0 \\ \pi_2 & \phi_2 & 0 & 1 \end{pmatrix} \cdots\cdots,$$

where ϕ, ψ, π are the component powers at a single surface, and the separations between successive surfaces are represented by $\mu\tau$. The gain secured by separating the wave curvatures from the constructional particulars of the system becomes more pronounced as the system grows in complexity. The discovery of an optical system to satisfy given conditions reduces itself to the routine solution of a matrix equation.

Space will not permit me to deal in detail with more advanced instruction. I would merely express my conviction that the line of progress lies in the application of Hamilton's methods to these problems. We effect a great simplification at the very beginning. We need not consider the form of the light-waves at all; the fact that they are absent from our fundamental matrix equations shows that they are not in fact pertinent to the problem. In Hamilton's methods such superfluous quantities are not introduced.

On the experimental side I have some ground for thinking that there is a widespread failure in places of higher education to realise what is reasonable accuracy in ordinary optical experiments. The special technique of the most refined measurements is a quite distinct matter, and does not fall within the province of educational institutions. To take a simple example it should be realised that for ordinary industrial purposes the refractive index of a glass is unknown unless its value is obtained to within two or three units of the fourth decimal place; the dispersion should be known to the same number of units in the fifth place. Greater attention should be devoted to the significance of the exact procedure followed in experimental observations. Accuracy of focussing for instance should be judged by sharpness, not by parallax. I suggest that some consideration should be given to the value, for instructional purposes, of a general utility interferometer. It does not appear to me that the cost of an efficient instrument would necessarily prove the decisive factor.

2. SUGGESTIONS RELATING TO REFORM IN THE TEACHING OF GEOMETRICAL OPTICS

By C. W. HANSEL, Bedford School

ABSTRACT. Reform in teaching will only come through pressure brought to bear by examinations. Experimental work should be devised in which the student must exercise his originality and resource in devising his apparatus and methods. The student should be given far more opportunity of making himself familiar with the construction and performance of optical instruments. Graphical exercises could be devised to supplement the experimental work. An appeal is made to mathematicians to state their problems wherever possible in non-mathematical language.

The type of training provided by our schools is largely adapted to suit the requirements of examinations. For this reason it may be said that examining bodies are almost entirely responsible not only for the subject-matter of what is taught but also for the method of teaching it. Any reform in the teaching of geometrical optics in schools can only come by modification of the syllabuses of the examinations for scholarships, university entrance, and school leaving examinations with simultaneous modification in the view-point of the examiners. Teachers, text-book writers, courses of practical work, and the equipment of school laboratories, will all adjust themselves to new circumstances and new ideas in order to conform to examination requirements; but reform, even if obviously urgent and desirable, will never come into the schools unless it is encouraged through the avenue of examinations.

To influence examining bodies to the extent of introducing fundamental changes of syllabus, will require very definite proposals of compelling urgency and with obvious educational advantages. Such a scheme can only be formulated by close and expert collaboration between teachers and industrialists. Men of knowledge and experience must cooperate sympathetically and must be given suitable facilities so that contact may be established between educational and industrial points of view. Sufficient time and labour must be devoted to the problem to allow of a thorough exchange of views. Any scheme formulated must take into account existing conditions and must provide some guidance for examiners and teachers. A report carefully drawn up and explaining the scheme, together with suggestions and exercises, could probably be published so as to be available at small cost.

It is possible that present methods of teaching may be too academic and non-technical. More attention might be given to industrial processes, e.g. optical glass, silvering, polishing, etc. Photometric work might be brought more into line with industrial requirements. More attention might be devoted to the construction and performance of optical instruments, the defects of optical systems and their cor-

rection, optical tests, etc. It is suggested that the methods employed in teaching geometrical optics are unsuitable and useless for dealing with industrial problems. It must be remembered, however, that beginners have a very meagre range of mathematical equipment, and that the majority of those in schools will never pursue their optical studies for industrial ends.

The most important part of the work should be the laboratory work. There seems to be a general tendency to regard practical work in optics as a series of experiments in which refractive indices, focal lengths, magnifying powers, etc. are measured in as many different ways as possible. A sufficient number of pin, parallax, and screen experiments would seem to provide the elementary student with the necessary optical training. Much that is useful can be done by such a course if it is suitably enlarged with graphical and numerical exercises, but there is danger of this work becoming a mere mechanical examination, routine. Too often the student is supplied with apparatus which suggests or even prescribes for him his method of experimenting and leaves nothing to his originality or resource. It is highly desirable that the student should exercise himself on the design of his apparatus and in devising satisfactory means for the solution of his problem. In the interests of economy, apparatus must be inexpensive but adaptable, and it should be capable of easy and universal adjustment in so far as this is required.

Perhaps the most important defect of present-day teaching is that it fails to supply a sufficient practical acquaintance with optical instruments and their performance. Many students of optics use telescopes and microscopes and they are given a simple theory of these instruments. Too often, they are quite unfamiliar with the construction of an eyepiece, the position and function of the lenses and stops in an instrument, questions relating to suitable illumination, etc. The construction of optical apparatus and instruments is a part of the practical training which is most seriously neglected. Apparatus and instruments should be capable of assembly in more than one way, and the discovery of the best method of construction and arrangement, and a criticism of inferior designs, is an essential and valuable part of any course of practical training. Those engaged in the optical industries would render valuable service if they would furnish some simple examples of preliminary exercises, numerical and practical, forming a suitable introduction to the design of optical instruments. Also, they could indicate some important tests, such as the Hartmann test, which might be modified suitably to enable them to be carried out effectively with simple apparatus.

The following apparatus is suitable for beginners. Let a number of parallel lines be ruled on an ordinary blackened lantern slide about $\frac{3}{16}$ in. apart. Let the slide be fixed in a vertical position with a bright flash lamp behind it and a cylindrical lens in front of it. By adjusting the position of the lamp, it is possible to project a parallel, convergent, or divergent beam of light on a white card placed in front of the slide. A second cylindrical lens may be placed in the path of the beam, and a test made of its spherical aberration. The chromatic aberration may be dealt with, using coloured glasses. The path of a ray through various optical elements may be exhibited by the use of a black slide with one or more slits.

Some brief reference may be made to elementary numerical and graphical work. In dealing with simple problems on axial pencils, I have found the following method simple and effective.

The figure shows the ordinary method of finding the position of the image formed by a thin convex lens where the position of the object is given. By rotation of the line PQ about O its intersections with $m = 1$ and the line QF give corresponding positions of object and image respectively.

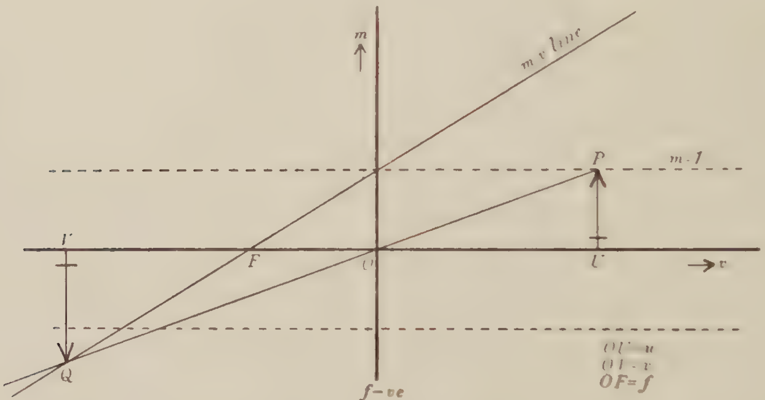
For a thin convex lens the magnification m is given by

$$m = -\frac{v}{f} + 1.$$

Thus the line QF is also the line obtained by plotting m against v.

The same diagram may be used to give a geometrical proof of the lens formula.

The magnifying power of the lens is the ordinate to the m, v line for $v = d$, where d is the nearest distance of distinct vision.



The method is quite general and may be applied to axial pencils through any optical system consisting of reflecting or refracting surfaces.

By the previous method the numerical and graphical treatment may be combined.

A graphical construction which I have found very useful in many problems (caustic curves, theory of rainbow, thick lenses, etc.) is given in Drude's *Optics**, and stated to be the "method of Weierstrass." It was given by Thomas Young in 1807 and by Weierstrass in 1858.

Attention may also be directed to a paper on graphical methods by Mr J. H. Dowell†. The methods here given are simple, elegant, and of wide generality. They are capable of numerical interpretation with the use of simple mathematics. Much valuable work could be done, by the methods described in this paper, with a large drawing-board, compasses, and parallel rule (or set squares).

* Drude, *Optics*, English translation, p. 32.

† J. H. Dowell, *Proceedings of the Optical Convention*, p. 965 (1926).

The *Proceedings of the Optical Convention* and the *Dictionary of Applied Physics* contain a mass of valuable information on optical problems. Much of this work carries little conviction to non-mathematical readers owing to the principles involved being totally obscured by a mass of mathematical symbols. The writer is of the opinion that many of these very valuable articles would make a much wider appeal if the mathematical results were stated alternatively in simple language and given a definite physical interpretation whenever this is possible. As examples of non-mathematical exposition of the kind advocated mention may be made of the papers contributed to the *Proceedings of the Optical Convention* 1926 by Miss H. G. Conrady* and Mr R. Kingslake†.

* Miss H. G. Conrady, "Spherical Aberration on Optical Images," *Proc. Optical Convention*, p. 830 (1926).

† R. Kingslake, "Developments of the Hartmann Test," *Proc. Optical Convention*, p. 839 (1926).

3. THE PRESENT SCHOOL CERTIFICATE COURSE IN OPTICS

By S. R. HUMBY, M.A., The College, Winchester

ABSTRACT. Many pupils after taking the usual first school course in Optics are found to have learnt little more than a geometrical construction and an algebraic formula. Emphasis should be placed on the artificial nature of the conventional treatment as well as on its usefulness. A course of very simple practical work, carefully designed to illustrate how optical instruments behave and how light passes through them, should provide this emphasis.

THERE still seems to be a strong tendency to regard the early course in optics as almost entirely a drill in geometrical diagrams and in elementary algebra. It is, incidentally, a matter of interest that the reciprocal equations are very difficult to many boys chiefly, it would seem, through lack of earlier practice in mental arithmetic and the conversion of simple fractions into decimals and decimals into their nearest simple fractions.

This treatment of elementary optics from a geometrical rather than from a physical point of view is of course partly due to the syllabus and to the type of question set in the First School Examinations.

It is comparatively easy for an examiner to set a variety of numerical examples on lenses or mirrors, and such questions are so much easier to mark rapidly and fairly that they are sure to hold an important position in the examinations and therefore in the teaching of the subject. As a result the writer has found that comparatively few of the abler candidates attempt the more qualitative questions on physical optics while the smaller allowance of time given to the other parts of the subject leave the weaker candidates hopelessly vague about them.

In School Certificate Examination scripts the answers show much confusion about the simple geometrical method of locating an image by a diagram. Many candidates have apparently never realised that the essential condition for a good lens or mirror is that all rays which strike it from a point on the object shall pass through a point image. It is evidently very necessary for us to emphasise the artificial nature of the conventional diagram as well as its great use.

The position in space of an image is another thing about which ideas are usually very vague. The writer, from his own experience, finds it best to begin by pointing out the two possible ways of fixing the position of a ghost by offering it a chair (if friendly) or by taking bearings on it from suitable positions. The location of images from mirrors, lenses and more complicated instruments can then be referred to these two possible methods.

In answers to questions which require the tracing of a pencil of rays through a system from a point on the object to the eye most candidates (and some text-

books) just give two geometrical diagrams for locating the image. It is evident that the pupils have never been made to realise that a ray of light which suddenly and abruptly changes its direction at a point in space is a most unnatural phenomenon. The trouble is mainly due, no doubt, to the difficulty of making clear diagrams in which the angles subtended by objects are reasonably small.

It seems that it is very necessary to provide practice in drawing approximate paths of rays which strike mirrors and lenses in directions other than the two or three standard ones. Very simple apparatus can be made to show these paths and those students who perform, or are shown repeatedly, simple experiments with rays of light are more likely to realise something of the relative amounts of light which pass into and are reflected back from a refracting surface. They will therefore recognise total internal reflection as a natural phenomenon rather than an apparently arbitrary mathematical relation.

In practical work the writer finds that much confusion is produced in the most elementary stages by the practice of finding focal lengths by four or five different methods, probably on different days. The pupils not unnaturally get the idea that these methods are arbitrary devices to defeat examiners. Perhaps the best way out of this difficulty is to insist on an approximate direct measurement of focal length by the distant object method each time before a new method is attempted. It seems also necessary to give a good deal of practice with real images caught on screens before attempting the much more difficult pin methods.

It is curious that unit magnification methods are so rarely described in elementary text-books for conditions where image and object can be made to coincide. Their equality in size is at once evident and, with beginners, adjustment for equality in length is a matter of seconds whereas the no-parallax position for a pin is very difficult to find.

4. REMARKS ON THE TEACHING OF GEOMETRICAL OPTICS TO ELEMENTARY STUDENTS

By H. H. EMSLEY

ABSTRACT. It is suggested that the teaching of the earlier stages of geometrical optics from the experimental point of view, the introduction of mathematics being postponed until later, would benefit larger numbers of students and at the same time provide those who are to pass on to more advanced work with a useful practical groundwork. In deciding upon a system of nomenclature and conventions, consideration should again be given to non-mathematical students and to the conventions established in the ophthalmic industry.

I REGRET that pressure of work has prevented me from joining very actively in what is undoubtedly a discussion of great importance to all interested in the teaching of optics. The points I am able to advance concern only the elementary portions of the subject of geometrical optics and amount practically to a plea that, in any scheme that may be formulated as a result of this discussion, consideration be shown to the non-mathematical student in general and to the large number of students of ophthalmic optics in particular.

Useful work in the adjusting, testing and using of optical appliances and even in the designing of many optical systems can undoubtedly be carried out by people equipped with little mathematics; and the teaching of geometrical optics in its earlier stages would benefit larger numbers of students if the subject were presented from the experimental rather than from the mathematical point of view. At present there is a wide gap between courses of study or text-books concerned merely with the simple elements of paraxial image formation, etc. (designed only too often merely to train for examinations and almost useless in practice) and courses or text-books on geometrical theory in which the whole presentation is supported on a mathematical framework which the non-mathematical student is unable to comprehend, although he may be quite capable on the experimental side and sound in his grasp of the physical nature of the problems involved.

I would suggest that the subject should be presented experimentally until the higher branches are reached. The students are learning mathematics simultaneously and those who are to proceed with the higher branches of geometrical theory will benefit by their early training on the experimental and practical side. In the elementary portions too, the paraxial (Gaussian) image formation should be developed as a special case after a certain amount of work has been done in tracing wide aperture rays and waves through surfaces and lenses, graphically, trigonometrically and experimentally, as is the case when dealing with prisms. For the experiments a large lens of about 10 or 12 in. aperture, capable of rotation about a vertical axis which may be in or beyond the plane of the lens, is extremely useful

in giving the student an idea of the meaning of chromatic and spherical aberration, coma, astigmatism, etc., diaphragms with small circular and annular apertures in various zones being used with the lens.

Although I would not advocate a rigid adherence to the curvature method of teaching geometrical optics, yet the method is helpful to the elementary student and can be usefully introduced alongside the ray method, capital letters being used to designate curvatures and powers and length reciprocals generally (see below).

The prismatic action of lenses should be presented. Experiments on the astigmatic image formation by sphero-cylindrical lenses could with advantage be introduced at an early stage, to give point to the meaning of stigmatic or homocentric image formation assumed in the geometrical theory when dealing with spherical surfaces.

Elementary ideas on interference and diffraction and the resolving power, numerical aperture, etc., of instruments might be incorporated much earlier than is usual.

The first ten chapters of Müller-Pouillet's section on optics, simplified for the beginner and supplemented by experimental, workshop and graphical work, seems to me to provide a useful guide in formulating a course on elementary geometrical optics, intended equally for those students who need to go no farther and for those who are to specialise on higher work in optics.

Great benefit would result from a few periods in the lens workshop devoted to the working and polishing of a simple lens surface on glass the refractive index of which has been measured by the students, the finished lens being measured for focal length on the optical bench.

The question of nomenclature and conventions as to signs, etc., is of minor interest in this discussion, I understand; nevertheless, having in mind the large numbers of ophthalmic students and the by no means negligible effect they will have on the optical industry, I would urge that any system of nomenclature that may be recommended should take into account the conventions that have become firmly established in the ophthalmic industry and the methods that have proved satisfactory in teaching students of ophthalmic optics. The main points may be summarised thus:

(1) A converging system has a positive focal length measured from the system to the second focal point. This fixes the direction of the incident light as the positive direction.

It should be an easy matter to fit this requirement into a wider and more rigorous scheme required by students of the higher branches of the subject*.

(2) The usual convention as to signs taught in schools should be adopted; distances to the right of and upwards from the point of reference to be positive, and (a point of less importance) counterclockwise rotations to be positive.

The incident light would therefore be taken usually as from left to right.

This would obviate the need for unlearning the notion of signs acquired by the elementary student in his school work, besides being practical.

* Müller-Pouillet, 3rd ed., chap. 3.

(3) Capital letters should be reserved for quantities the dimensions of which are L^{-1}—curvatures, vergencies, powers of systems, etc. To avoid confusion these capitals could be in script form as distinct from the block capitals used for other purposes. Whether the curvature method of teaching geometrical optics be used or not, the use of separate symbols for reciprocals of lengths such as curvatures and focal powers is undoubtedly helpful to the elementary student. It was urged by the late Mr Chalmers.

Thus the power of a system (in air) of focal length f is denoted by F (or $\mathscr{F}$); the curvature of the wave-front of light incident from a distance u is represented by U (or $\mathscr{U}$); the curvature of a surface of radius r by R (or $\mathscr{R}$), etc.

These three underlying conventions are already in use at the Northampton Institute and, I believe, with one or two minor exceptions, at the Imperial College of Science. The symbols used at the former institution in designating the more important points, distances, etc. have recently been published in another place.

5. GEOMETRICAL OPTICS AT THE CAVENDISH LABORATORY

By Dr G. F. C. SEARLE, F.R.S., University Lecturer in Experimental Physics, Cambridge

ABSTRACT. The paper is based on experience with a practical class at the Cavendish Laboratory. Many students show little care for accuracy. They are encouraged to attempt accurate measurements: this arouses their interest. The optical parts supplied to students should be of good quality. Students should be encouraged to use a magnifying glass in making accurate settings. The convention of signs adopted is that which makes all signs positive in the formula for a converging lens. Thick lenses and lens systems are taken at an early stage in the course.

I am not qualified to speak of the teaching of geometrical optics generally, for my limited experience has been gained in my own practical class at the Cavendish Laboratory, and in that work I have had little occasion to teach more than the elements of the geometrical theory.

At Cambridge, geometrical optics has a small place in Part I of the Mathematical Tripos, but the questions can be answered without any practical acquaintance with the subject and not many of those who teach for that Tripos have themselves any practical knowledge of optical work.

Optics was removed from Schedule A of Part II (the advanced part) of the Mathematical Tripos some years ago on the ground, I believe, that it was not intellectually interesting, and thermodynamics was put in its place. The effect, of course, has been to diminish the number of serious mathematicians who know anything about optics and to deprive teachers of the opportunity of teaching the subject.

Some years ago Mr T. G. Bedford and I gave some experimental lectures on geometrical optics for candidates for Part I of the Mathematical Tripos. We did not attempt a mathematical treatment but merely endeavoured to show the simpler phenomena on a bold scale and to indicate how measurements can be made. The students were obviously interested, but the lectures were discontinued after the second year because pressure on the lecture-room accommodation prevented us from arranging the experiments.

At the Cavendish Laboratory, both geometrical and physical optics, to use a common distinction, are taught to students preparing for Part I of the Natural Sciences Tripos. The practical work in geometrical optics includes experiments on mirrors, prisms and lenses; the range of work may be seen in my *Experimental Optics*.

In Part II of the Natural Sciences Tripos, the optics is "physical" and there is no insistence on geometrical optics.

The teaching for Part I of the Natural Sciences Tripos is intended more as a

part of a general education than as providing a specialised training in any one subject, and thus, if we can impart a sound knowledge of the fundamental principles of geometrical optics and can teach students how to do accurate and methodical work in experiments which interest them, we have fairly fulfilled the duty allotted to us.

In a course of lectures on geometrical optics, the lecturer is free to select a number of propositions and to arrange them in the best possible way. But, in a practical class, the experiments are determined by the apparatus which the teacher has at his disposal. Thus, experiments on the variation of refractive index with wave-length may have to wait until a mercury arc lamp is available. Since only a limited number of experiments can be done in one term, the more exciting experiments are apt to crowd out the less exciting. This process is often aided by the students who have, many of them, become blasé after several years of school physics. If an experiment buzzes or stands on its head or is otherwise thrilling, students will clamour for it, though they may understand little of the principles involved. For this reason experiments on geometrical optics are in danger of being left on one side.

In my own class we have to deal with students who have done physics for as many as five or six years in some cases. Many have never been taught to care for accuracy and are content to continue in their old way. The majority are quite happy in focussing a telescope on a distant object through dirty crinkly window glass, and only a few have any idea how to find an approximate value of the focal length of a converging lens by observing the image of the sun. Many take little care in placing the object point on the axis of the lens. They are awarded the O.N.E.—the Order of the Near Enough.

> Oh, let me be dismal and work in the dark
> With lenses all crooked and dim,

seems to express their outlook.

We do our best to cure this state of things and our efforts meet with fair success. I think that the cause of the trouble is that the students have not been sufficiently encouraged to attempt accurate measurements. Geometrical optics, done anyhow, is very dull. When, however, the experiment is hard enough to call forth a real effort, the student "gets his blood up" and finds interest and pleasure in the work. One student told me that he took no interest in optics till I made him work accurately.

I am fully aware that the geometrical optics taught in my class would need much development if the training of practical opticians were in view. But to teach students to recognise small differences in the quality of images and to lead them to appreciate the importance of accuracy in aligning lenses is not generally possible when one teacher is expected to teach about fifteen students and when, as sometimes happens in universities, the junior teachers have little more experience than the students. I do not know that much encouragement is given to teachers to make themselves proficient. If a young man is engaged on Research, he is a Pioneer and a Hero, but if he teaches—well, he is just a teacher.

I would make an earnest plea that the optical parts supplied to students should be of good quality. Teachers should not be content to provide a piece of shiny material that is not obviously uneven for use as a plane mirror. It will not be possible to provide optical flats for students, but if pieces of stout mirror glass are selected by aid of a telescope, students are unlikely to meet with any difficulty arising from the imperfection of the mirror. If possible, students and laboratory assistants should be taught how to test a plate against the face of a good plate or prism by means of Newton's rings.

My eyes have for some years lost so much power of accommodation that I have forgotten what it is like to see images with the eyes of a young man and I have to use a magnifying lens in cases where 40 years ago I should not have needed its aid. But I feel sure that students should be taught at an early stage to use a magnifying lens in making optical settings in order that they may get the stimulus which comes from an endeavour to work accurately.

We often find that a student will focus an eyepiece so that the image of the cross-wires formed by the eyepiece is not at infinity but is perhaps not more than 10 cm. from his eye, although his vision for distant objects is not in the least defective. It is useless to expect such a student to use a microscope or a telescope to determine the position of an object by adjusting the microscope or telescope along an optical bench, unless the instrument be provided with cross-wires or with a micrometer scale which may be used in the parallax test.

Some students stare at an image for a long time before passing any judgment on its position. This spoils their vision and leads to gross errors. I have known an able student make errors of 10 or 20 cm. on an optical bench in experiments where a normal person would not be wrong by $\frac{1}{2}$ cm. unless he were very careless. Such students need to learn to come to a speedy decision.

Some students have not learned to focus accurately. They will work with a spectrometer and a grating capable of giving a very obvious resolution of the *D* lines of sodium and will not see the lines separate until the teacher helps them. On the other hand, when the laboratory possesses good spectrometers, the teacher is always pleased to hear a student ask what is the purpose of the vertical wire down the centre of the slit, for then he knows that the student has, unawares, discovered the separation of the *D* lines for himself.

In one respect I have departed from some text books. I call the focal length of a converging lens positive and, when the real object and its image are on opposite sides of the lens, I call both u and v positive, so that the equation becomes

$$1/u + 1/v = 1/f.$$

I believe I have been blamed for this although I have several respectable authorities on my side.

When the object is a flame or other luminous source and the image is received on a screen, there may be some point in measuring both distances "towards the light." But when the object is one pin point and a second pin point is set on the image of the first, there seems no reason for taking one direction as positive rather

than the other. It is therefore more natural and more simple to measure the distance of the object from the lens in one direction and the distance of the image in the opposite direction and thus gain the benefit of a symmetrical notation for a symmetrical arrangement.

As regards taking the focal length of a converging lens as positive, this is exactly what is done by practical opticians.

When a person skilled in optical calculations has to deal with a system of many lenses, he may find it more convenient to measure all distances in the same direction. He will escape disaster because he is skilled and will not make mistakes as to the signs of the quantities. But, when an ordinary person has to deal with a simple lens, he does best to adopt the simplest method and that method is the symmetrical method. My experience leads me to think that among unskilled persons those who make the most conventions make the most mistakes. Some years ago a physicist told me that he never could tell by calculation on which side of a lens the image is, and I had no reason to doubt his word.

I am not sure that the "teaching of geometrical optics" amounts to much more than talking about the eternal question as to whether to write $1/u + 1/v$ or $1/u - 1/v$. To have to decide this point is for many a worse trial than the attempt to find out when the twentieth century began.

An examination of the books in which the focal length of a converging lens is taken as negative and distances are measured "towards the light," and a consideration of the explanations given to make it all perfectly simple, will convince the candid reader that there is something wrong.

When a student arrives in my class with his mind entirely fogged as to sign conventions, I teach him the deviation method of treating thin lenses and then his difficulties, as a rule, disappear. All he needs to remember is that, if a ray in a plane containing the axis meet a thin lens at distance h from the axis, the direction of the ray is changed by h/f in passing through the lens.

We do not treat a thin lens as the standard form of lens and then consider it a hardship to be compelled to treat some lenses as thick, but we begin at an early stage with thick lenses or systems of lenses, using exact formulae. When the fundamental theorems are approached by way of Helmholtz's formula they present little difficulty to students.

I feel sure that much of the confusion over $1/u + 1/v$ and $1/u - 1/v$ arises from a failure to distinguish between the algebraical sign prefixed to a symbol in an equation and the positive or negative character of the numerical value which is assigned to the symbol when an actual numerical computation has to be made. This trouble is not likely to distress students who have already a sound knowledge of the fundamentals of algebra and are sure of what they do with symbols. But many who do optical experiments have not got that sound knowledge. They can do algebra up to a point and no doubt would suspect an error if the number of bulls came out to be "$-$ 14·027," though they might explain the negative sign as indicating cows. But when in optics they find by calculation that the image is 17 cm. from the lens on the same side as the object there is nothing to warn them that they are wrong

unless they make an experimental test or follow the advice of some text books and verify their calculation by aid of a graphical construction. Systems of calculation which need graphical confirmation stand self-condemned.

Optics suffers from a special disadvantage which limits the appreciation of the science by non-technical teachers and students. An image of sorts can be obtained with a lens which is not "corrected" for the purpose in hand and with very poor setting, and this image is sufficiently good to satisfy many. It is difficult to excite their interest in the technical skill involved in the design of lenses to fulfil definite conditions, in the methods of constructing lenses or in the art of producing the optical glasses of which the lenses are made. Few have ever seen the grinding of lenses or the figuring of prisms or mirrors, and fewer still have done these things with their own hands, and thus their lack of appreciation of the work of technical opticians is only natural. Besides all this, optics is an old science and therefore does not appeal to those who hope to make names for themselves in the latest developments, probably ephemeral, of physics.

Those, if any there be, who look to me to give a prescription for putting geometrical optics into a better position will be disappointed. The elements of the subject are simple and are easily taught, but their applications to practical optics, when fully developed, are far from simple, as may be gathered from such a book as the *Principles and Methods of Geometrical Optics*, by James P. C. Southall, or from any book on lens design. It is therefore not to be expected that there will ever be many who are expert on the theoretical side.

If each teacher presents geometrical optics in a convincing and interesting way within the limitations under which he works, this is perhaps as much as teachers can do for the subject at present. Greater prosperity in the optical trade would provide opportunities for the development of the subject, but this aspect of the matter is outside my province.

6. THE TEACHING OF OPTICS BY THE WAVE OR CURVATURE METHOD

By C. V. DRYSDALE, O.B.E., D.Sc., F.R.S.E.

ABSTRACT. The teaching of optics by the wave or curvature method is advocated on the grounds that it (*a*) simplifies elementary teaching, (*b*) harmonises the teaching of science students and opticians, and (*c*) leads naturally to higher physical optics. A large number of opticians present themselves annually for examinations, and the majority of them find it difficult to obtain suitable instruction, as the trade nomenclature is based on the curvature system. Although it is generally known that wave methods can be employed for elementary teaching, there seems to be an impression that they must be abandoned for the study of aberrations and optical design, and that it is therefore better to start with geometrical methods at the outset; but this is a mistake. The first order aberrations can be more easily dealt with by the curvature treatment, and the results harmonise completely with the indications of the lens testing interferometer. The paper contains a sketch of a systematic course on optics, including the Gauss theory and the treatment of aberrations, in which wave methods are employed.

§ 1. INTRODUCTION

ALTHOUGH it is fairly generally known that the whole of elementary optical theory can be taught by the wave or curvature method, and with considerable simplification in many parts of the subject, few teachers appear to have adopted it, and probably the late Prof. Silvanus Thompson* and the present writer are alone in having employed it systematically.

To some extent this may be due to the lack of a complete text-book on the subject, but there seems to be a general impression that the wave method is unsuitable for the treatment of aberrations, and that since the geometrical or ray method must be employed for this purpose it is better to adopt it from the outset. This impression is however incorrect, and in the writer's optical lectures at the Northampton Institute, not only was the wave method used for the whole of elementary optics, but it was carried on to the teaching of aberrations and the design of lens systems. An outline of the elementary course was given to the Physical Society in 1905†, and a fairly complete exposition of it in a series of articles in the *British Optical Journal* ‡, but no publication of the application to aberrations was made until the Traill Taylor lecture before the Royal Photographic Society last year§.

* "Notes on Geometrical Optics," S. P. Thompson, *Phil. Mag.* **33** (1), 232–248 (1889).

† "On the Curvature Method of Teaching Optics," *Proc. Phys. Soc.* **19** (1905). *Phil. Mag.* April (1905).

‡ "Applied Optics," *British Optical Journal*, (1901–3).

§ *The Photographic Journal*, March, April, and May (1928).

The advantages of the wave method of teaching may be stated as follows:

1. Elementary optics is considerably simplified, as trigonometrical formulae and approximations are very rarely required. Signs and conventions are also simplified.
2. The gap between scientific optics and spectacle optics is removed.
3. It puts the whole of optical teaching on a physical basis, and leads naturally to the study of interference, diffraction, and polarisation.

The second point is of considerable importance. At the present time, some hundreds of opticians present themselves annually for the examinations of the Spectacle Makers' Company and the British Optical Association, and it is extremely difficult for these candidates to obtain satisfactory instruction, owing to the small number of teachers who have given attention to spectacle optics. It would greatly increase the efficiency of these opticians if the teaching of optics in all the science centres were suited to their needs, and it would also be advantageous if all those who studied optics were familiar with the language and methods employed by the opticians. Unless, therefore, there is good reason for separating scientific teaching from the trade methods, it is highly desirable that the teaching in all the centres should be suitable for opticians. Indeed, as the methods of opticians are in complete harmony with physical optics, it is difficult to justify the separation.

In view of the long time which has elapsed since this question was last discussed, it appears desirable to recapitulate briefly the main features of the wave treatment; referring those specially interested to the papers above cited.

§ 2. ELEMENTARY OPTICS

Nature and propagation of light. The propagation of transverse waves is shown by a ripple tank, which may advantageously be illuminated stroboscopically so that the apparent rate of propagation may be slowed down. The Huygens explanation of approximately rectilinear propagation can be well shown by employing multiple sources and increasing their number until they become a practically linear source. Elementary ideas on interference diffraction and polarisation may usefully be introduced at the outset; and the pinhole camera serves as an introduction to magnification in optical systems.

Illumination and photometry. Each wave contains a certain amount of light energy, and the energy per unit area of wave surface or illumination therefore diminishes as the square of its radius. Oblique illumination and the cosine law follow from the projected area of the surface.

Reflection at plane surfaces. Huygens's construction for plane waves. Position and size of image follows from symmetry of wave fronts before and after reflection.

Curvature theory and notation. Curvature R is defined as the angle turned through per unit length of arc, and proved to be the reciprocal of the radius r, or $R = 1/r$. The unit of curvature is the dioptre or curve of 1 metre radius, so that

$$R_{\text{dioptres}} = 1/r_{\text{metres}} = 100/r_{\text{cm}} = 39{\cdot}37/r_{\text{inches}}.$$

It is desirable to exhibit a curve of 1 metre radius, so that this curvature may

be visualised and to point out that for a chord of 9 cm. (more accurately 8·95 cm.) the curvature in dioptres is represented by the sag in millimetres.

Measurement of curvature, spherometer. In any circle by Euclid III, 35,

$$(2r - s)\, s = c^2,$$

where s is the sagitta or "sag" at the centre of a chord of length $2c$. Hence

$$R = 1/r = 2s/(c^2 + s^2).$$

If $\qquad s < c/10, \quad R = (2/c^2)\, s$ within 1 per cent.,

so that for a given length of chord the curvature is proportional to the sag. If R is to be in dioptres s and c must be in metres, but if they are expressed in centimetres $R = (200/c^2)\, s$. It should be emphasised at this point that the approximation is only justified for short chords or small apertures, and that as the chord or aperture is increased the term h^2 becomes more important, the way being thus paved for the later study of aberrations. The ordinary and direct reading dioptre spherometers are then explained.

The curvature theory and notation apply equally to reflecting or refracting surfaces and to wave fronts, and the most convenient convention for the latter is to consider all convergent waves as of positive curvature, and to speak of that curvature as the convergence in dioptres, irrespective of the direction of travel.

Reflection at curved surfaces. If a plane wave impinges on a concave mirror the sag of the wave front after reflection is evidently double that of the mirror, so that the convergence $F = 2R$ where R is the curvature of the mirror. If the incident wave has convergence U or $1/u$, the distance between the wave front and mirror for a given chord is proportional after reflection to $U + R$, which $= V - R$. Hence

$$V - U = 2R = F,$$

corresponding to the ordinary relation for conjugate face

$$1/v - 1/u = 2/r = 1/f,$$

except for the change of sign of V. A concave mirror always has positive, and a convex mirror negative convergence. Magnification may be treated in the ordinary manner, but the most simple general method is to consider the mirror stopped down to a pinhole at the vertex, in which case

$$m = v/u = U/V,$$

from the pinhole camera relation. The effects of reflection at concave surfaces may be particularly well illustrated with the ripple tank, stroboscopic illumination being used, as if the frequency of the flashes is a little less than that of the waves there is an apparent slow progress from the source to the conjugate focus, while if the frequency is increased the waves appear to travel in the opposite direction, clearly showing the conjugate principle.

Refraction at plane surfaces. The velocity of light being assumed to be reduced in dense media, Huygens's proof of the sine law

$$\sin\theta/\sin\phi = V_1/V_2 = \mu_2/\mu_1 = \mu$$

is given, and may be illustrated by the ripple tank, and by the idea of ranks of men

marching obliquely towards and crossing a river. If these ranks are made up of tall men in red uniforms and short men in blue, the reason for dispersion is clearly seen. Thick prisms are treated in the ordinary manner, but the relations

$$\theta_1 + \theta_2 = \delta + \alpha, \text{ and } \phi_1 + \phi_2 = \alpha$$

are more easily seen when wave fronts are used instead of rays. Thin prisms are particularly simple: since while the wave traverses a thickness t of glass it would traverse a distance μt in air, it is retarded by the distance $(\mu - 1)\,t$ so that

$$\delta = (\mu - 1)\,\alpha,$$

where δ is the deviation and α the angle of the prism. The optician generally expresses the deviation of thin prisms in "prism dioptres" or the number of centimetres displacement in a metre, so that one prism dioptre, 1Δ, is 0·01 radian or 0·573 degree.

When spherical waves pass from one medium to another we evidently have the relation

$$\mu_1 s_1 = \mu_2 s_2 \text{ or } \mu_1 U = \mu_2 V,$$

so that

$$V = (\mu_1/\mu_2)\,U = U/\mu.$$

This gives the "apparent thickness" of a block of glass, and the "microscope method" of measuring refractive indices. The quantity μs may be called the "optical sag," and all problems of refraction may be dealt with by the principle that when waves pass from one medium to another the optical sags of the incident and refracted wave fronts are equal.

Refraction by thin spherical lenses. Since the retardation of a wave in passing through a thickness t of medium of refractive index μ is $(\mu - 1)\,t$, it follows immediately that in passing through a lens the change of convergence

$$V - U = (\mu - 1)\,C = (\mu - 1)\,(R_1 + R_2) = F,$$

where R_1 and R_2 are the curvatures of the two faces and C or $(R_1 + R_2)$ is the "total curvature" of the lens. This is equivalent to the ordinary formula

$$1/u - 1/v = (\mu - 1)\,(1/r_1 - 1/r_2) = 1/f,$$

but with reversal of sign of $1/r_2$, the reason being that in the case of a double convex lens both surfaces converge the light and are therefore reckoned as positive. The relation

$$F = (\mu - 1)\,(R_1 + R_2)$$

may be written

$$F = (\mu - 1)\,R_1 + (\mu - 1)\,R_2 = F_1 + F_2,$$

where F_1 and F_2 are the "surface powers" of the lens, which may be measured directly by a "Geneva lens measurer." This instrument is a direct reading spherometer having two fixed points and a third sliding point midway between them, which point is geared to a pointer moving round a scale and is pushed outwards by a spring. When the three points are pressed against a plane surface the pointer registers zero on the scale, but if the surface is convex or concave it indicates the "surface power" (positive or negative) on the assumption that the index of refrac-

tion of the glass is about 1·52—the ordinary value for spectacle glass. Two fixed points are used instead of the three of the ordinary spherometer, in order that it may be possible to measure cylindrical or toric lenses. In measuring thin spherical lenses the instrument is applied to both faces, and the power of the lens is the algebraical sum of the readings.

The optician is concerned with the form of lenses as well as their power, as meniscus lenses are generally preferable for spectacles. The above treatment shows at once that the form of a lens of any given power may be varied to any extent by pushing in one face and pulling out the other, so long as the central thickness or total curvature remains the same. The optician is also concerned with cylindrical, sphero-cylindrical, and toric lenses and their transposition, and these subjects can be most simply dealt with on the above lines. The magnification of the image is obtained as with mirrors by supposing the lens to be stopped down to a pinhole, giving the same relation

$$m = v/u = U/V.$$

Decentration. Another important matter is the combination of lenses with prisms to correct defects of convergence in the eyes. A convex lens of convergence F or $100/f_{cm}$ dioptres brings all axial parallel light to a focus on its axis at a distance of f_{cm}, so that a ray at a distance d_{cm} from the axis is deviated d_{cm} in f_{cm} or $100d/f$ or Fd_{cm} in a metre, which is the prismatic effect Δ in prism dioptres. Hence

$$\Delta = Fd, \text{ or } d = \Delta/F,$$

so that the decentration in centimetres required to produce a given prismatic effect is obtained by dividing the number of prism dioptres required by the power of the lens in dioptres. The base of the prism is in the same direction as the decentration with convergent lenses, but in the opposite direction with divergent lenses.

Thin lenses in contact. Neutralisation. In the dioptric notation, the power of any number of thin lenses in contact is obviously the algebraic sum of their individual powers; provided, of course, that the total thickness of the combination is small. This fact, in combination with the prismatic effect of decentration, is the basis of the opticians' method of measuring the powers of lenses by "neutralisation." The lens to be tested is held before the eye and some distant object and is moved vertically or horizontally. The object will appear to move in the same direction as the motion of the lens if the lens is convergent, but in the opposite direction if it is divergent. Known lenses of opposite power are then taken from a trial case and combined with the lens to be tested until the apparent motion is neutralised.

Effect of axial displacement. The convergence of a spherical wave obviously increases as it travels towards its focus. If its initial convergence is F dioptres and it travels forward a distance d metres, its convergence will be

$$F' = 1/(f - d) = F/(1 - Fd)$$

dioptres. This is of some importance in spectacle optics, owing to the varying displacement of lenses from the eye; and it forms an introduction to the theory of separated lenses, thick lenses, and lens combinations.

§ 3. LENS COMBINATIONS

Separated thin lenses. If we have two lenses of powers F_1 and F_2 dioptres separated by a distance d metres, and parallel light falls axially on the first, the convergence of the light reaching the second lens will be U_2 or $F_1/(1 - F_1 d)$ dioptres, and

$$V_2 = F_1/(1 - F_1 d) + F_2 = (F_1 + F_2 - F_1 F_2 d)/(1 - F_1 d)$$

dioptres after passing through it. This corresponds to the ordinary formula for the "back focus," $(f_1 + f_2 - d)/\{f_2(f_1 - d)\}$. If we have a distant object subtending an angle α, the first lens would form an image of length $f_1 \alpha$ or α/F_1, and this will be magnified by the amount m_2 where

$$m_2 = U_2/V_2 = F_1/(F_1 + F_2 - F_1 F_2 d)$$

by the second lens, giving the size of the final image as

$$\alpha/(F_1 + F_2 - F_1 F_2 d).$$

But if we have a single lens of power F dioptres the size of the image will be α/F, and if this is to be equal to that given by the combination,

$$F = (F_1 + F_2 - F_1 F_2 d),$$

which is therefore the equivalent power of the combination, corresponding to the formula $f = f_1 f_2/(f_1 + f_2 - d)$ for the equivalent focal length.

Refraction at a single spherical surface. In dealing with thin lenses we considered the effect of the lens as a whole, without troubling about the separate surfaces, but for thick lenses and combinations the refraction at each surface must be considered separately. If we have a convex surface of curvature R_0 between two media of refractive indices μ_{-1} and μ_{+1} then, by the principle of equality of the optical sags,

$$\mu_1 (V_1 - R_0) = \mu_{-1} (U_{-1} - R_0), \text{ or } \mu_1 V_1 - \mu_{-1} U_{-1} = (\mu_1 - \mu_{-1}) R_0 = F_0.$$

Also if we consider a pinhole aperture at the vertex,

$$m = v \tan \phi/u \tan \theta \doteq v \sin \phi/u \sin \theta = \mu_{-1} v_1/\mu_1 u_{-1} = \mu_{-1} U_{-1}/\mu_1 V_1.$$

Using von Seidel's notation we may denote $\mu_{-1} U_{-1}$ by U'_{-1}, the "reduced" convergence, and $\mu_1 V_1$ by V_1', from which we have

$$V_1' - U'_{-1} = (\mu_1 - \mu_{-1}) R_0 = F_0, \text{ and } m = U'_{-1}/V_1'.$$

Thick lenses and lens combinations. If we have a thick lens of curvatures R_0 and R_2 and thickness d_1, then at the first surface

$$V_1' - U'_{-1} = F_0 \text{ and } m_0 = U'_{-1}/V_1$$

as above. In traversing the lens the convergence V_1 will increase to

$$U_1 = V_1/(1 - V_1 d_1),$$

from which

$$U_1' = \mu_1 U_1 = \mu_1 V_1/(1 - V_1 d_1) = \mu_1 V_1/\{1 - \mu_1 V_1 (d_1/\mu_1)\} = V_1'/(1 - V_1' d_1'),$$

if $d_1' = d_1/\mu_1$. But d_1/μ_1 is the "apparent" thickness of the lens so that if we employ

"reduced" convergences and "apparent" thicknesses, the formulae for thick lenses are identical with those for separated combinations of thin lenses.

From the three formulae

$$V_+' - U_-' = (\mu_+ - \mu_-)\,R = F,\ \ m = V_+/U_-,\ \text{ and }\ U_+' = (V_+'/(1 - V_+'d'),$$

we can obtain the Gauss formulae for lens combinations and the positions of the principal, focal, and symmetrical planes, and nodal points, as described in the Physical Society paper above cited.

§ 4. ABERRATIONS

Only a short sketch of the leading points of aberration theory can be given here, but a fairly complete theory has been given in the Traill Taylor lecture above cited.

Chromatic aberration. Since for a thin lens $F = (\mu - 1)\,C$,

$$\Delta F = C\Delta\mu = F.\Delta\mu/(\mu - 1) = F/\nu,$$

where ν is the "efficiency" of the glass and

$$\nu = (\mu - 1)/\Delta\mu.$$

If we have two thin lenses in contact

$$F = F_1 + F_2,\ \text{ or }\ \Delta F = \Delta F_1 + \Delta F_2 = F_1/\nu_1 + F_2/\nu_2.$$

Hence for an achromatic combination

$$F_1 + F_2 = F,\ \text{ and }\ F_1/\nu_1 + F_2/\nu_2 = 0,$$

from which

$$F_1 = \frac{\nu_1}{\nu_1 - \nu_2}F,\ \text{ and }\ F_2 = -\frac{\nu_2}{\nu_1 - \nu_2}F.$$

Spherical aberration. In dealing with the theory of curvature, it was pointed out that the sag was only proportional to the square of the chord for small apertures. To determine it for large apertures suppose a convergent wave U falls on a surface of curvature R. Then in the annexed diagram, Fig. 1, the sag

$$s = u - u',\ \text{ and }\ u'^2 = r^2 + (u - r)^2 + 2r\,(u - r)\cos\alpha.$$

If h is the oblique distance from the point considered to the vertex

$$\sin\alpha/2 = h/2r,\ \text{ and }\ \cos\alpha = 1 - h^2/2r^2,$$

from which

$$u'^2 = u^2 - h^2\,(u - r)/r$$

and

$$s = u - \sqrt{u^2 - h^2\,(u - r)/r} = u\,\{1 - \sqrt{1 - U\,(R - U)\,h^2}\}.$$

Expanding, we have

$$s = (R - U)\,h^2/2 + U\,(R - U)^2\,h^4/8 + U^2\,(R - U)^3\,h^6/16 + \text{etc.}$$

Hence the "optical sag"

$$s_1' = \mu_1 s_1 = \mu_1\,(R - U)\,h^2/2 + \mu_1 U\,(R - U)^2\,h^4/8 + \mu_1 U^2\,(R - U)^3\,h^6/16 + \text{etc.}$$

If we now assume for the moment that the wave front after refraction is spherical and of curvature V, then by the same reasoning

$$s_2' = \mu_2 s_2 = \mu_2 (R - V)\, h^2/2 + \mu_2 V (R - V)^2\, h^4/8 + \mu_2 V^2 (R - V)^3\, h^6/16 + \text{etc.}$$

For this to be true $s_2' = s_1'$ or $s_2' - s_1' = 0$, but from the above

$$\begin{aligned}\Delta s' = s_2' - s_1' = &\{\mu_2 (R - V)^2 - \mu_1 (R - U)\}\, h^2/2 \\ &+ \{\mu_2 V (R - V)^2 - \mu_1 U (R - U)^2\}\, h^4/8 \\ &+ \{\mu_2 V^2 (R - V)^3 - \mu_1 U^2 (R - U)^3\}\, h^6/16 + \text{etc.}\end{aligned}$$

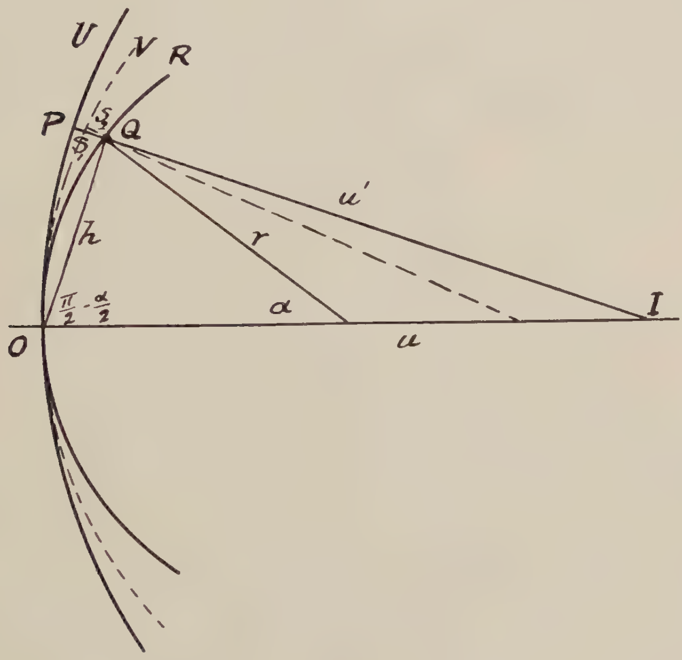


Fig. 1.

The first term vanishes when

$$\mu_2 V - \mu_1 U = V' - U' = (\mu_2 - \mu_1)\, R = F_1,$$

the ordinary relation for small apertures, but this does not make the higher terms vanish, and we therefore have:

First order spherical aberration

$$\Delta s' = \{\mu_2 V (R - V)^2 - \mu_1 U (R - U)^2\}\, h^4/8.$$

Second order spherical aberration

$$\Delta s' = \{\mu_2 V^2 (R - V)^3 - \mu_1 U^2 (R - U)^3\}\, h^6/16, \text{ etc.}$$

Where $\Delta s'$ in each case is the departure of the wave front from the spherical form, measured normally to the sphere in optical distance or wave lengths. It should be noticed that this exactly corresponds with the measurement of aberrations by the Hilger lens testing interferometer, and that it enables the interferograms

obtained with this instrument to be directly interpreted. Objection has been taken to this method, owing to the complicated form which the distorted wave front takes as it advances towards the "focus"; but if the aberration is small this is not of importance, while if it is large, it must be corrected before it has advanced very far, so that the objection has no practical weight. In dealing with the spherical aberration of a lens system, it is only necessary to trace the course of a pencil approximately through the system, and to calculate $\Delta s'$ at each surface; the total displacement of the wave front from the spherical form after emergence being the algebraic sum of the displacements at the several surfaces.

Oblique aberrations. Before dealing with these generally, it is useful to familiarise students with the general character of such aberrations by taking the simple case of a single thin convergent lens. It was shown above that the power of such a lens

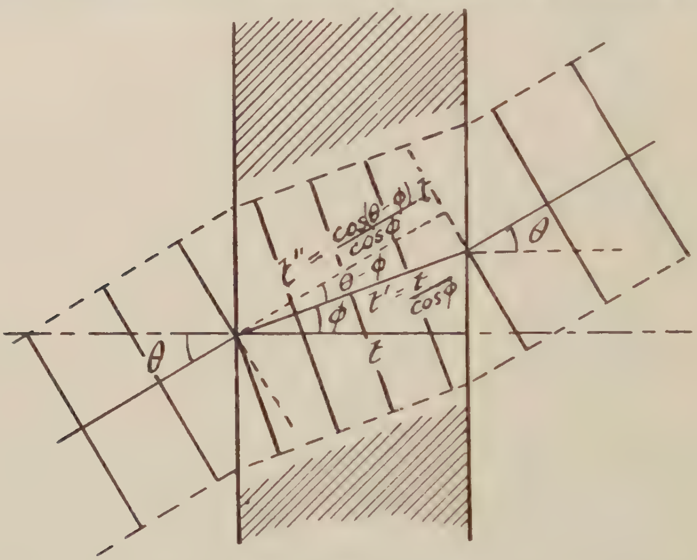


Fig. 2.

could be directly obtained by considering the retardation of the centre of the wave front on passing through the glass. If we have a parallel plate of glass of thickness t as in Fig. 2 and consider plane waves falling on one surface inclined at an angle θ, the actual distance traversed through the glass is $t \cos \phi$ where ϕ is the inclination of the refracted wave to the surface, and in the time taken to traverse this distance the wave would have advanced by $\mu t \cos \phi$ in air. The actual distance travelled by the wave front will however be $(t \cos \phi) \cos (\theta - \phi)$ owing to the change of inclination, and the retardation is therefore

$$\mu t/\cos \phi - t \cos (\theta - \phi)/\cos \phi = (\mu \cos \phi - \cos \theta)\, t,$$

as compared with $(\mu - 1)\, t$ for axial light. The ratio of the oblique to the axial retardation is therefore

$$(\mu \cos \phi - \cos \theta)/(\mu - 1) = 1 + \theta^2/2\mu,$$

when θ is small. Applying this to the thin lens, we find that the power in the sagittal

plane is increased in the ratio $1 + \theta^2/2\mu$. But in the meridional plane the obliquity of the wave front causes the same retardation to be given to a beam which is narrower in the ratio $\cos\theta$, so that the curvature is increased in the ratio $1/\cos^2\theta = 1 + \theta^2$, when θ is small. Hence we have:

Power in meridional plane $= F(1 + \theta^2/2\mu)(1 + \theta^2) = F\{1 + \theta^2(2\mu + 1)/2\mu\}$.

Power in sagittal plane $= F(1 + \theta^2/2\mu)$.

The lens therefore acts as a sphero-cylindrical lens of spherical power $F(1 + \theta^2/2\mu)$ and cylindrical power $F\theta^2$, the axis of the cylinder being parallel to the axis about which it is tilted—a result which is useful to the spectacle optician. It is easy from this to work out the curvature of the primary and secondary focal surfaces, and to give clear ideas of astigmatism and curvature of the field; and as the wave front meets the lens asymmetrically, some idea of coma can be formed. This simple treatment strikingly exhibits the advantages of the wave method, as no other method enables the idea and magnitude of the oblique aberrations to be so easily demonstrated.

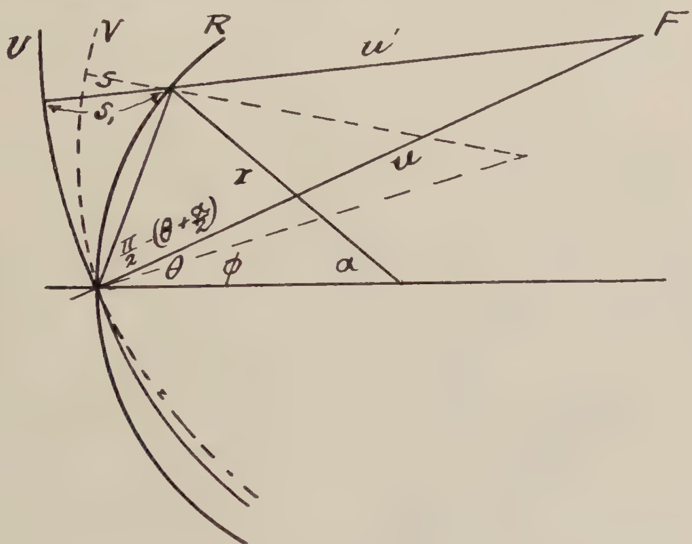


Fig. 3.

Oblique refraction at a spherical surface. After the above introduction and the treatment of axial spherical aberration, the calculation for oblique refraction at a single surface is a relatively easy matter. In the annexed diagram, Fig. 3, we have the same conditions as in Fig. 1 except that the axis of the incident wave front U is inclined at an angle θ to the radius of the refracting surface at the point of incidence. We then have:

$$u'^2 = u^2 + h^2 - 2uh\cos\{\pi/2 - (\theta + \alpha/2)\} = u^2 - \{2u\sin(\theta + \alpha/2) - h\}h,$$

and expanding as before, and putting

$$\sin\alpha/2 = Rh/2, \text{ and } \cos\alpha/2 = 1 - R^2h^2/8 - R^4h^4/128, \text{ etc.}$$

we have

$$s_1 = \sin\theta . h + (R\cos\theta - U\cos^2\theta)\, h^2/2 + \{U(R\cos\theta - U\cos^2\theta) - R^2/4\} \sin\theta\, h^3/2$$
$$+ \{U\,[R\cos\theta - U\cos^2\theta\,\{R\cos\theta - U\,(1 - 5\sin^2\theta)\} - R^2\sin^2\theta]\}\, h^4/8 + \text{etc.}$$

which reduces to the same expression as that obtained for the axial pencil when $\theta = 0$.

If, as before, we assume that the refracted wave front is circular and of curvature V we have

$$\Delta s' = s_2' - s_1' = \mu_2 s_2 - \mu_1 s_1 = (\mu_2 \sin\phi - \mu_1 \sin\theta)\, h$$
$$+ \{\mu_2\,(R\cos\phi - V\cos^2\phi) - \mu_1\,(R\cos\theta - U\cos^2\theta)\}\, h^2/2$$
$$+ [\mu_2\,\{V\,(R\cos\phi - V\cos^2\phi) - R^2/4\} \sin\phi$$
$$- \mu_1\,\{U\,(R\cos\theta - V\cos^2\theta) - R^2/4\} \sin\theta]\, h^3/2 + \text{etc.}$$

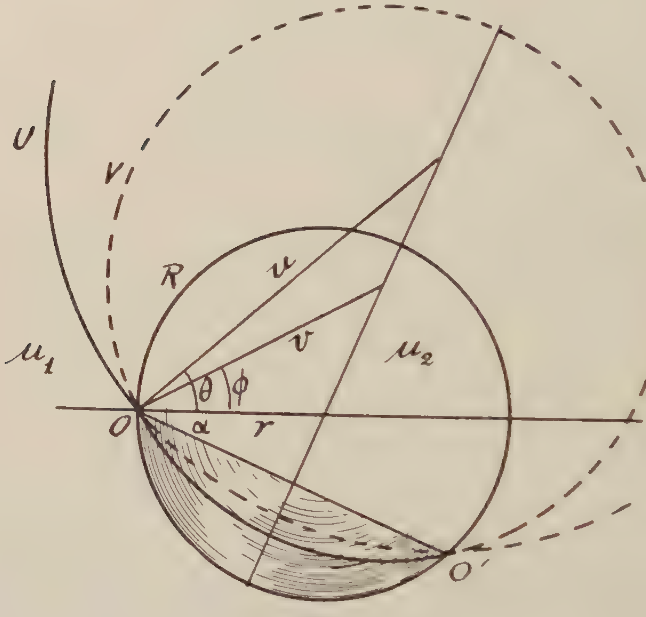


Fig. 4.

The first term vanishes when $\mu_2 \sin\phi = \mu_1 \sin\theta_1$, the ordinary law of refraction, and the second vanishes when

$$\mu_2 V\cos^2\phi - \mu_1 U\cos^2\theta = V'\cos^2\phi - U'\cos^2\theta = (\mu_2\cos\phi - \mu_1\cos\theta)\, R;$$

leaving the third term, which is obviously asymmetrical, to give the optical displacement of the wave front due to coma.

The above calculation applies only to the meridional plane. For the sagittal plane a similar process may be adopted, but for the first approximation a more simple method may be employed. In Fig. 4 the whole of the spherical surface R is shown, with the two wave fronts U and V intersecting it at the points O and O'.

If we consider the sphere cut through by a plane OO', the circular section will have a radius $r \cos \alpha$, and considering the wave fronts as spheres we have evidently

$$u \cos (\theta + \alpha) = v \cos (\phi + \alpha) = r \cos \alpha,$$

from which

$$\cos \theta - \sin \theta \tan \alpha = r/u = U/R \quad \text{and} \quad \cos \phi - \sin \phi \tan \alpha = V/R,$$

and since

$$\mu_1 \sin \theta = \mu_2 \sin \phi,$$

$$\mu_2 V - \mu_1 U = V' - U' = (\mu_2 \cos \phi - \mu_1 \cos \theta) R$$

for the sagittal plane, as compared with

$$\mu_2 V \cos^2 \phi - \mu_1 U \cos^2 \theta = V' \cos^2 \phi - U' \cos^2 \theta = (\mu_2 \cos \phi - \mu_1 \cos \theta) R$$

for the meridional plane. It is easy to show that these results agree with those obtained from the simple treatment of the thin lens.

The expression for coma becomes

$$\Delta s' = \{V (R \cos \phi - V \cos^2 \phi) - U (R \cos \theta - U \cos^2 \theta)\} \mu_1 \sin \theta \, h^3/2.$$

Distortion may be considered to the first approximation as spherical aberration of the exit pupil, and therefore requires no fresh treatment.

In the author's own lectures, a method of tracing and summing the aberrations through a system was developed, analogous to the method of von Seidel; but this has not been published, and it would involve too much space to be given here. Enough has, however, been said above to show that the wave or curvature method can be employed with advantage for the treatment at least of first order aberrations, and the harmony of this method with the indications of the lens testing interferometer is a feature which must commend it. For the final design of large aperture systems, ray tracing is undoubtedly necessary, but this presents no difficulty to students who have followed the wave method to this stage, and hardly justifies the idea that geometrical methods should be adopted at an early stage to provide for this particular application. The diffraction phenomena at and near the focus can of course be readily comprehended from the form of the wave front.

It is unnecessary to carry the treatment further into the field of physical optics in the present paper as the whole method has been physical from the outset and there is therefore no discontinuity to be bridged. Brief and imperfect as this sketch has necessarily been, it is hoped that it will convince many teachers of the practicability and advantages of systematic optical teaching by the wave method, and that an increasing number will adopt it, with the result of making their teaching suitable for science students and opticians alike.

7. FIRST STEPS; THE PROBLEM OF THE BEGINNER

By V. T. SAUNDERS, M.A.

ABSTRACT. The general conditions which a school course in geometrical optics must satisfy are considered, and a method of presentation which has been suitable for teaching beginners is outlined. It is pointed out that there is no time available during school life for a serious course in geometrical optics beyond the investigation of the ideal lens, and also that even if there were time the boys have not sufficient mathematical equipment to handle the problems involved. The suggestion is made that more attention might be paid in the initial stages to the simpler conceptions of physical optics, more especially for those boys who will not specialise in science.

SCIENCE masters in schools must approach this problem as learners. We have to survey the whole accumulated body of knowledge of light and decide how boys may be introduced to it; after a selection of material suitable to form an introduction we have to consider how it may best be presented in order that a true appreciation of the principles of the subject may be grasped.

In making our selection of material and formulating our method of presentation we have to be guided by our knowledge that it is only a small percentage of our boys who will become professional scientists, and that the vast majority will follow other callings in life; they will be the people who will avail themselves of the facilities which science provides to assist them in their lives, but they will not make any contribution to the advancement of knowledge in the abstract, nor to the improvement of technical methods. Again we have to bear in mind that youth is more concerned to do than to think but that the faculty of thinking can be stimulated by interest in doing. It follows from these considerations that a school science course must satisfy two conditions: (i) it must be complete in itself as far as it goes, dealing with general principles and widespread applications; (ii) it must be a sound foundation on which those who carry on subsequent study, may build. In this present discussion we are concerned to discover if this second condition is, in fact, satisfied by the courses at present followed in schools.

A school course must begin with the concrete and work towards the abstract. The sunbeam offers us a concrete starting point; it becomes a ray and its path can be followed. Straight line propagation and parallax are accepted and the study of lenses can be started. We give a boy a lens and lead him to discover what it will do. I find by experience that this method can be followed without any discussion of the sine law of refraction. The image-forming property is arresting, and its various aspects are soon discovered and appreciated. Combinations of two lenses follow easily, and a simple telescope and microscope are soon made and understood. We work from the parallel ray definition of the principal focus, and from the magnification property; the formulae involving reciprocals are avoided, and all

quantitative work is done by drawing ray diagrams by the collinear method. Plane and curved mirrors are investigated in the same way. When boys handle lenses, microscopes and telescopes, they learn the language of the subject, and also the potentialities and limitations of the instruments; familiarity with things that are leads them to ponder over what might be. This initial stage must be largely experimental; the demonstration experiments can be well shown with a piece of apparatus known as the "optical disc" (which is supplied by a number of apparatus firms) and by the smoke box. The laboratory experiments for the boys can be carried out with apparatus which costs very little.

This method of treatment may be retained up to the First School Examination, but it appears that the method of procedure to be followed by those boys who stay at school up to the age of 18–19 years and go to a university to read science, is the main question under discussion. At present the general method followed in schools, in teaching geometrical optics, is to accept the spherical surface of refraction and to ignore the defects of the images formed by it. This undoubtedly leads to failure to appreciate the actual powers and limitations of optical instruments. On the other hand we in schools must teach general principles and the extent to which we can follow out the technique of geometrical optics is strictly limited. It can be argued with much justification that the educational value of physical optics might be explored more fully, and that, for a general education, knowledge of the simpler aspects of physical optics gives a broader outlook on the field and method of science than is provided by an intensive study of the technical knowledge involved in the design of optical instruments.

However, after the First School Examination stage a number of important points have to be faced: (i) the sign convention; (ii) the introduction of the wave conception; (iii) the differences between the ideal lens and the actual lens. With regard to the first, I find that after boys have been started on ray diagrams they find little or no difficulty with the sign convention which accepts as positive, distances measured against the direction of the incident light; boys using the reciprocal formulae almost invariably draw a freehand diagram to aid them in fixing their signs. I have considered the other, and sometimes simpler, conventions for the initial stages, but I am unable to see that any advantage is gained in not accepting the convention which is at present found in all advanced text-books. If, in the future, the methods and conventions of the university text-books change, it will, I think, be necessary to modify our methods (in this second stage) in order that uniformity may be reached.

On the question of the introduction of the wave conception, I find that it comes most easily after a discussion of the velocity of light. The development of the reciprocal formulae from wave-fronts, for small apertures, involves no difficulties at this stage. The properties of waves and the effect of change of the velocity of propagation of waves can be adequately demonstrated by a ripple trough illuminated by a point-o-lite lamp. The simpler experiments showing interference fringes can be demonstrated with advantage, and an estimation of the wave-length of light arrived at.

On the question of the actual, as opposed to the ideal lens, very little is taught in schools: the mathematical equipment necessary for the solution of the problems involved is usually not to be found in schoolboys. But if a simple method of presenting this branch of the subject is disclosed it will certainly find its way into school work.

I have tried to point out what is being done in schools at present, and to give reasons for our methods. No attempt is made to justify the present procedure; it remains for those who consider that our methods are not the best, to indicate how we may improve; the school science master must keep an open mind on these matters, and modify his methods to conform with the progress of knowledge.

Finally I would plead for the retention of optics as a school subject. It is sometimes said that because we cannot teach boys the more exact methods and processes of the technical aspect of the subject, it is better not to teach it at all. I believe that this attitude is based on a misconception of the function of school education, and that it is detrimental to the progress of science. The school education gives a boy an insight into the things that are being done in the world, it enables him to be in sympathy with the work carried out by researchers, and it supplies the best means by which the general body of public opinion may learn to accept scientific research and its applications to industry, as essential factors for the continuance and for the development of the present civilisation.

8. AN ELEMENTARY COURSE ON LIGHT BASED ON THE WAVE THEORY

C. G. VERNON, Bedales

ABSTRACT. The writer describes his experience in teaching light from the beginning from the point of view of the wave theory, a method which he states to be both possible and easy. Starting from a consideration of waves, illustrated by means of wave-machines and a ripple-tank, he develops Huyghens' principle and investigates thoroughly the phenomena of propagation in straight lines, plane reflection and refraction, and the simple theory of the function of glass blocks, lenses and spherical mirrors. Besides giving a good grounding in these fundamentals, the method leads naturally to the study of the spectrum and the more advanced phenomena. The writer indicates the development of the mathematical side with the use of curvatures, and suggests that the correct rôle of geometrical optics is to serve for the revision and coordination of work already done. He indicates difficulties that arise over sign conventions, and states that this is a distinct drawback to the teaching. Finally he makes a plea for the carrying out of practical work by beginners with real beams of light, as distinct from the misleading and uninteresting methods known as "pin and parallax."

BEING dissatisfied with the accepted methods of teaching elementary light I devised a method of approaching the subject from the point of view of the wave theory. This was taught to pupils who had completed two years of physics in a secondary school course, and proved to be highly successful. The following account tells of my objections to the older methods together with the advantages I consider this method to possess, an outline of the elementary part of the course and of the mathematical development of it, the very real difficulties that exist over sign conventions, some criticisms of the course and a reference to methods of demonstration and practical work.

In drawing up the course I derived considerable help from the late Silvanus Thompson's book, *Light, Visible and Invisible*. I have embodied my experience in a text-book that has recently been published by the Cambridge University Press, and which contains an amplification of the points set out below.

§ 1. OBJECTIONS TO ORTHODOX METHODS

1. Lack of interest. The lessons tend to become sterile and lifeless.

2. The theory is almost entirely mathematical, and has little or no reference to the physical nature of light.

3. The reasoning involved is restricted in nature, and there is no scope for the exercise of controlled imagination, which is, or should be, one of the essentials of the teaching of physics in schools.

§ 2. ADVANTAGES OF THE NEWER METHOD

1. Interest.

2. The physical nature of light is made the basis of a very thorough grounding in the essentials of: propagation; plane reflection; plane refraction (and total reflection); function of lenses and mirrors; colour; relation to radiant heat.

3. The reasoning is varied, and so makes a wider appeal than does that of the other method.

4. Light ceases to be in a water-tight compartment, but takes its place as a branch of physics, related to heat, properties of matter, electricity and modern views on the constitution of matter.

5. The necessary introduction to the study of waves is a valuable thing in itself, and also serves as an introduction to sound.

§ 3. OUTLINE OF THE COURSE

(*a*) Nature of wave-motion; simple harmonic motion. Distinction between motion of wave and movement of particles of the medium. Transverse and compression waves.

Wave length; frequency; velocity of wave; amplitude.

(*b*) Ripples on water (ripple tank). Disturbance from one point and simultaneously from many points in a line, giving circular and straight-line wave-fronts.

Effect of isolating a small portion of the wave-front. Huyghens' principle and constructions based on it.

(*c*) Use of Huyghens' principle to investigate the persistence of shape of a wave in a uniform medium. Concave wave-fronts and the existence of a focus.

(*d*) Plane reflection of straight-line and circular waves (1) from the construction, (2) as seen in the tank. Effect of introducing a sheet of glass into the tank and so producing a second (shallower) medium differing from the first. Refraction of straight-line and circular waves due to a difference in velocity.

(*e*) Applications to light. Straight-line and circular ripples correspond to plane and spherical waves in space. Effect of the increasing area of a spherical wave — inverse square law and simple photometer. Parallel, diverging and converging beams of light.

Plane reflection and position of the image.

(*f*) The "refraction roller" experiment to illustrate refraction of a plane wave. The velocity constant, a physical constant of the medium.

Huyghens' construction used for refraction of a plane wave. Apparent depth of water due to refraction of a spherical wave. Concept of a ray as "a line representing the direction of movement of an element of the wave-front." Use of rays for simplification of diagrams.

N.B. Avoid referring to "a ray of light."

(*g*) Shadows. (Mention difficulty about shadows and waves.) Umbra, penumbra, eclipses, etc.

(*h*) Simple refraction problems. Parallel-sided glass block. Semicircular glass block, critical angle and total reflection. In each case investigate (1) with roller, (2) by Huyghens' construction, (3) with light, e.g. smoke box.

Mention Snell's law at this point, using μ as a mathematical ratio.

Other cases may then be dealt with, namely ghost images in a thick mirror, total reflection prism, fish-eye view, mirage, setting sun, etc.

(*i*) Effect of reflecting straight-line waves at a circular surface—imprinting of curvature.

Illustrate in a tank, by Huyghens' construction, and by means of beams of light on spherical mirrors in a smoke box.

(*j*) Functions of lenses, imprinting of curvature due to slowing down of either centre or periphery of a portion of wave-front. Types of lens. Illustrate by smoke box and draw both wave and ray diagrams.

As may be judged, the student may by now be assumed to be familiar with the fundamental phenomena of propagation, reflection and refraction and their consequences, and the mode of operation of lenses and spherical mirrors. The criticism may be levelled that it is not possible for the students to do much practical work that has any close bearing on the theory of matter, but the drawing of the wave diagrams is certainly just as much "practical light" as is the "fixing of rays by means of pins" (!); and, using narrow beams of light (such as may be obtained by means of the "Rugby" optical apparatus) the beginner may do the qualitative demonstrations of the behaviour of light that have been indicated above.

§ 4. MATHEMATICAL DEVELOPMENT

There are now two courses open. By the use of the wave concept it is possible to establish all the important relations by means of simple proofs that are easily followed by children of reasonable intelligence. A summary of the more important points of the method may be of interest.

Call the velocity-constant (i.e. velocity in air ÷ velocity in medium), h.

Use the definition, curvature = reciprocal of radius, and then establish the relation, sagitta = curvature × a constant.

Focal power, denoted by F, is taken as the curvature imparted to any wave by the given lens or mirror, and is measured by the curvature given to a plane wave (whence $F = 1/f$).

The following may be shown:

1. Position of image behind a plane mirror.

Sagittae of incident and reflected waves are equal but of opposite sign, whence

$$v = -u.$$

2. Relation between μ and h (velocity-constant),

$$h = \frac{\text{velocity of wave in air}}{\text{velocity in the medium}} = \frac{\sin r}{\sin i} = \frac{1}{\mu}.$$

3. Real and apparent depth.

The sagittae of the incident and refracted waves are in the ratio of $1/h$ ($= \mu$).

4. Position of focus of spherical mirror.

The sagitta of the reflected wave is twice that of the mirror surface, whence

$$V = 2R,$$

i.e.

$$1/v = 2/r,$$

and thus $v = r/2$, where v is the distance of the image from the plane wave (i.e. the focal length).

§ 5. UNIVERSAL FORMULAE FOR LENSES AND MIRRORS

This proof has manifest advantages over the one which makes use of certain rays. If the curvature of the incident wave from a distance u be called U, that of the emergent wave which will form an image at a distance v be called V, and the focal power of the lens be F, then

$$V = U + F.$$

Expressed as reciprocals this becomes

$$1/v = 1/u + 1/f,$$

whence

$$1/f = 1/v - 1/u.$$

In the case of mirrors, as in all reflections, the original curvature is reversed, and the focal curvature is then added:

$$V = -U + F,$$

whence

$$1/f = 1/v + 1/u.$$

The more advanced theory can be dealt with by similar means.

§ 6. THE PLACE FOR GEOMETRICAL OPTICS IN A SCHOOL COURSE

It is my considered opinion, based on my teaching experience, that not until the student has reached the point of understanding the principles of reflection and refraction, with applications, and the functions of lenses and mirrors, should geometrical optics be used. It then affords a most valuable revision of work done, serves to clarify ideas and is a help in relating the practical and theoretical work, but does not assume an importance that cannot be justified. The ray should be defined as "a line showing the direction of movement of an element of the wave-front," and it must be made quite clear that geometric lines of light have no existence.

§ 7. OTHER ITEMS OF THE CONTENT OF THE COURSE

The spectrum and colour. Complementary colours. Pigments. Colour matching and mixing. Achromatism. Sky colours. Rainbow. Invisible spectrum: ultra-violet, infra-red, radiant heat, "wireless."

Measurement of the velocity of light.

The eye and its structure. Peculiarities of normal and defective vision. Aids to vision. Optical instruments. Projection apparatus.

Sources of light. Efficiency. Measurement of intensity of light and illumination. Lumens and foot-candles.

Some reference to interference, diffraction and polarisation.

§ 8. DIFFICULTIES

These have proved less than might be anticipated. All pupils understood the basis of Huyghens' construction, which had been demonstrated by means of ripples. They enjoyed doing the diagrams. It is most important to let demonstrations with the tank and with light overlap, so that there shall be no danger of their being regarded as two unrelated series of phenomena.

With respect to the mathematics, the brighter pupils can cope with the "wave" proofs and enjoy doing them. For slower people it is probably wiser to limit one's teaching to the "geometric" proofs, which in any case should be given as alternatives. The main difficulty, and a serious one at that, arises over the question of sign convention.

§ 9. SIGN CONVENTION

The international convention of opticians agreed to call a converging lens positive and a diverging lens negative. Hence it is only reasonable to call the curvatures imprinted in each case by the corresponding sign. Thus light from an object reaching a lens is diverging and therefore negative, while light converging on a real focus is positive in sign. On reflection there is a reversal of curvature, and this is indicated by a change of sign.

On the other hand, in the teaching of geometrical optics, the only satisfactory method I have found is the one set out below, which follows naturally from the graph work done in algebra:

> Always draw a diagram and consider the lens or mirror as situated at the intersection of the axes of a graph. Always place the object on the right-hand side of the lens or mirror, when u becomes positive, and allot a positive or negative sign to v according to whether the image comes on the right or left of the "origin" (as in algebraic graphs).

Thus it is seen that exactly opposite signs are obtained according to the method used, and of course this is a serious hindrance to teaching. I know of no way of overcoming the difficulty unless the opticians can be persuaded to change their nomenclature.

§ 10. GENERAL REMARKS

I am convinced, from my own experience, of the immense superiority, from the general cultural aspect, of the wave-concept approach to the study of light. The lessons were a joy to conduct, giving obvious pleasure to the students, and constantly stimulating their interest. Moreover, the ideas brought to the fore are essentially physical, and this aspect of the subject is not swamped under a flood of mathematical detail.

One other point deserves mention. It is most important in doing class experiments on light to use light and not pins. The latter may be less troublesome, but the value of experiments with pins is almost negligible in many cases, for the brighter pupils anticipate the results and the weaker ones never reach any. When beams of light are used even the quick workers find sufficient to hold their interest, and the slower ones are able to realise that optics is not such an unfathomable mystery after all.

9. GEOMETRICAL OPTICS FOR BEGINNERS

By L. MOORE, M.Sc., Leeds Grammar School

ABSTRACT. While the wave theory affords a useful means of revising geometrical optics it is not a suitable starting point. A pupil should use the same sign conventions throughout his school career; later, when more adaptable, he will not find much difficulty in making a change if that is found to be necessary. Experiments in which the images of luminous sources are projected on to screens by means of mirrors and lenses are to be preferred to experiments by the unconvincing pin methods.

§ 1. THE WAVE THEORY AS A STARTING POINT

USUALLY not enough ground is covered in courses on elementary light, owing to the large amount of geometry introduced and to the numerous, but unconvincing, experiments by pin methods. There is a danger that the introduction of the wave theory at an early stage in the course might merely add alternative proofs of the mirror and lens formulae, and that even less ground might be covered than at present. If, however, the wave theory could be used to simplify and reduce the amount of geometry to be learnt and allow time to discuss interference, etc., the change would be welcome.

Prof. Silvanus P. Thompson adopted a method of teaching geometrical optics by means of the wave theory*. If the method had proved successful, it would probably be in use to-day for beginners in optics. My own opinion is that while the wave theory affords a useful means of revising geometrical optics at the end of the higher school certificate course it is not a suitable starting point.

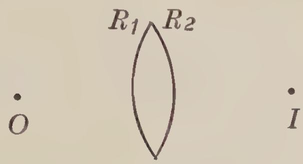
§ 2. SIGN CONVENTIONS

It is unfortunate that there are so many different sign conventions in use in text-books on light, and I believe that for the most part these are not the conventions used by opticians. Until I began to teach, the conventions I used were:

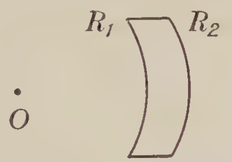
Concave mirror formula $\quad 1/u + 1/v = 1/f = 2/R$.

* *Phil. Mag.* **38** (1), 232 (1889).

Convex lens formula $1/u + 1/v = 1/f = (u - 1)(1/R_1 + 1/R_2)$.



The formulae and the diagrams go together: u, v, etc., are all positive. Changes in sign in the formulae are easily made.



$$1/u + 1/v = (u - 1)(-1/R_1 + 1/R_2),$$

where u, v, R_1, R_2, etc., are again all positive.

If v works out to be negative, then I is on the same side as O.

There was very little difficulty in changing over to another system of sign conventions to secure uniformity in the school, and I believe that if any particular sign conventions be decided upon and used throughout the school years, the few students who continue to study optics will not find much difficulty in making a change later, if that is found to be necessary, for they will then be more developed and more adaptable.

§ 3. EXPERIMENTAL WORK

A series of class experiments is needed using (i) luminous sources and projecting their images on screens by means of lenses and mirrors; (ii) reflection and refraction of beams of light instead of the usual tracing of rays by sighting pins. Young boys could then carry out experiments such as the following:

(1) Use a convex lens to project on a screen the image of the filament of an electric lamp. Without moving the screen find another position for the lens so that the image is focussed on the screen. Explain by means of two diagrams the difference in the size of the image in the two cases.

The answer to this practical problem would be two diagrams and nothing else.

(2) Try to project on a screen the image of the filament of an electric lamp using a concave lens. Draw a diagram to show why you cannot succeed.

10. THE PLACE OF GEOMETRICAL OPTICS IN THE TEACHING OF LIGHT IN SCHOOLS

By H. TUNLEY, Merchant Taylors' School, Liverpool

ABSTRACT. A scheme is outlined for the treatment of light up to university scholarship standard. In this, geometrical optics has its customary prominent place. An argument is presented against its being displaced by a wave theory treatment, while its limitations can be indicated even to beginners. It is suggested that in the treatment of curved mirrors and thin lenses there is need for much simplification and rationalisation, but a drastic departure from traditional methods does not seem advisable. There is a considerable need for modernising the treatment of photometry in general. Up-to-date knowledge is not readily available and most textbooks are either out of date or misleading on this branch. A suggestion is made concerning the further treatment of light for the "scholarship boy."

Is there anything very much wrong with geometrical optics if properly presented? Is not the trouble largely due to muddled presentation and confusion as between teacher and teacher, textbook and textbook?

The elementary teaching of light can be dealt with under four main groupings founded upon:

(1) Rectilinear propagation in a uniform medium.
(2) "Laws" of regular reflection.
(3) "Laws" of regular refraction.
(4) Inverse square law.

Here we are dealing with facts—limited in application though some of them may be—which can be satisfactorily demonstrated even to the youngest student. The building of the structure on these foundations is an exercise of logical deduction which has, surely, a considerable cultural value.

By using a Hartl optical disc (or a similar appliance), the behaviour of a narrow pencil of light can be visually demonstrated, and the young student feels that he is dealing with realities he can clearly conceive rays. Unless the student is older and has a good mathematical equipment before starting the subject, a wave theory presentation at the outset is not advisable. There is a difficulty in visualising a three-dimensional transverse wave, and by the time the teacher has been brought (by questions asked) to confess that "the ether may or may not exist but, anyway, we don't know anything about it and, moreover, these transverse waves are probably electro-magnetic in nature," the youthful mind is in a quandary and thinks it is dealing with unrealities. In the early stages a presentation based upon a theory is not as satisfactory as one based solely upon experimental facts and geometry.

If its limitations are carefully pointed out when opportunity arises, the method of geometrical optics is the more desirable treatment, and one, *par excellence*, in

which there will be nothing to unlearn later. If there is anything to unlearn it will be due to faulty presentation. Of the muddle that has arisen over the vexed question of signs we, in schools, are not entirely blameless, but first year university lecturers, and textbook writers are at least as much at fault. It is the experience of the author that geometrical optics can be presented without any great departure from the traditional methods, so that students have no difficulty over signs, and even on going through an honours physics course at a university, they find nothing to unlearn. This statement is made after due enquiries. Of the needs of the technical man the author is unaware, but is certain that, if by the time a student contemplates taking a post with a firm such as Ross's or Hilger's he is not independent of mere conventions and capable of quickly understanding a new outlook, then he is not able enough or "advanced" enough for the work. It is doubtful whether 1 in 1000 students of light in schools or 1 in 100 in the universities ever finds a career in technical optics, so that unless the technical man can expound to us a treatment of elementary optics as logical and simple as the present geometrical optics (properly presented) he must not expect from us more than a passing mention of his pet methods—even when we know what they are.

Appended are the mere outlines of a treatment of some parts of geometrical optics, but little of which may be original. It is not so much originality which is needed as careful systematic treatment. Inversion of an order or omission of an essential step makes all the difference.

APPENDIX

(1) *Rectilinear Propagation.* After shadows and eclipses have been satisfactorily dealt with the pin-hole camera is usually studied. This gives an opportunity to point out the limitations of rectilinear propagation. The pupil having suggested, as the result of "leading" questions, that decreasing the size of the hole will sharpen the image, it can be pointed out that there is a limit to this truth. This lesson can be followed up by the illustration, by experiment or photographs, of what happens to light from a narrow slit passing through a V-shaped slit. A suggestion to students to experiment at home with water ripples passing through variable gaps paves the way for an explanation by the wave theory later.

(2) "*Laws*" *of reflection.* Lecture demonstrations and measurements with an optical disc and individual experiments with apparatus such as Gallenkamp's optical ray apparatus, or that devised by Mr Meier of Rugby, will serve to convince the student of the truth of these so-called "laws."

There is much to be said for deducing all the well-known properties of plane and curved mirrors by simple mathematical steps from these laws and testing by experiment the results obtained at each step. This is rather a reversal of the method usually adopted in schools but it appeals to boys, if carefully done, and gives them confidence in the applications of mathematical methods to physics.

If the drawing of a caustic curve for the trace of a hemispherical surface precedes the usual study of spherical mirrors, the statement "the results hold only when the ratio *aperture/radius of curvature* is small" is fully comprehended.

One of the reasons why this part of the subject is often such a trial to young students is the mystification caused by all sorts of similar triangles, which will persist in getting mixed up, and mathematical tricks which often evade the best efforts of memorising. Then there is that vexed question of "signs." It is possible to present these things in a straightforward manner which appeals to the student without imposing any great strain on his memory. One simple illustrative example can be given in outline.

Point source on axis. Take any ray OA, reflected ray cuts axis at I (appeal to first "law" to show that it must cut axis). u is object distance, v image distance (mnemonic: u comes before v in alphabet, there must be an object before there can be an image). Establish connection between α, β and γ by using the facts that the exterior angle = sum of two interior opposite angles and $i = r$.

Then $\beta = AP/R$ exactly, $\alpha = AP/u$ approximately. (Draw actual arc of circle with OP as radius and show the extent of the approximation.) Similarly $\gamma \doteq AP/v$.

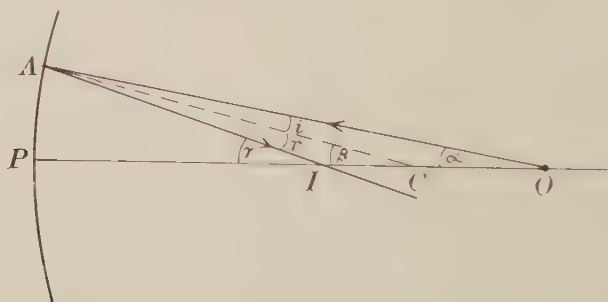
Thus prove that $\quad 1/v + 1/u = 2/R. \quad$ ......(1)

This is independent of the position of A, and therefore true for all rays from O within the limits of the approximation, $\therefore$ all cut at I. $\therefore$ I is image of O.

Take $OP < R/2$ and, working similarly, prove $-1/v + 1/u = 2/R$. ......(2)

Take convex surface and prove $-1/v + 1/u = -2/R$. ......(3)

Then argue that (1), (2) and (3) are very similar in form.



A convention of signs makes it possible, whilst remembering (1) only, to reproduce automatically the correct expression in any particular case. This excites interest (any labour or memory saving device invariably does!), and the convention is learned thus:

(*a*) Consider all distances as measured from the mirror (or lens) outwards.

(*b*) Assign a negative sign to those measured with the light and a positive to those measured in a direction contrary to that of propagation of the light, both when obtaining and when using a formula.

Thus when deducing (2) we shall prefix a negative sign to v (this is seen from a simple drawing) and thus find that the equation obtained now takes the form of (1). When however substituting a particular value for the variable v (as in the case we are discussing, where the image is on the negative side), we substitute

$v = -3$, say, which clearly brings the equation back to the right form for (2). Thus the student is convinced of the reasonableness and usefulness of the convention.

The author has also found it possible to simplify and rationalise the treatment of magnification, etc. without departing greatly from tradition.

As most higher schools certificate students know something of the properties of conic sections, the parabolic mirror should not be omitted.

(3) *Refraction.* Matters of apparent depth and total internal reflection present no difficulty, initially, to the ray treatment. Colour effects are noticed in the experimental work. Refraction through a prism leads to the observation that μ is different for different colours and that the prism decomposes white light. Clearly, this cannot be deduced directly from the refraction "laws"—it is well just to hint at this stage that it can be explained on a wave theory, but to deal with it as an experimental fact.

Refraction through a lens can be treated via the thin prism or may be taken up after a preliminary treatment of refraction at a curved surface. It is of educational value to do both. Refraction at a curved surface should be treated in any case and the method should be similar to that for reflection using the approximation $i = \mu r$ instead of $i = r$. This, then, presents no new difficulty to the student. Aplanatic foci should be emphasised. In proceeding in this way to the lens formulae the approximations introduced relative to aperture and thickness are made clear. The same sign convention has the same meaning and the same usefulness. After some general acquaintance with the properties of lenses the student has no difficulty in understanding that $1/f$ measures the "power" of a lens or in realising why the technical man regards the converging lens as having positive power although the convention makes its focal length negative. He realises that positive and negative in these cases refer to two totally different conceptions—power to converge a beam in the one case and spatial position of the principal focus in the other.

Students find the sign convention a servant and not a hard task-master—it is often more useful and less confusing than some of the ingenious schemes for avoiding it.

The "thin prism" method of approach, which should certainly be given to higher school certificate students in addition to the above, yields, at an intermediate step (at least by Dr Searle's well-known method), the important relation that $\delta = h(\mu - 1)(1/r_2 - 1/r_1)$, where δ is the deviation, h is the perpendicular distance from the axis at which the ray passes through the thin lens and the sign convention is used. This leads to a possible method for tracing the path of a ray through a system of thin lenses.

At this stage the following point of view, for which the author is indebted to Assoc. Prof. J. Rice, has interest and value:

$1/f$ measures the power of the lens to cause convergence.
$1/u$ measures the divergence of the incident beam.
(Divergence of inc. beam) − (convergence due to lens)   (remaining divergence),
i.e. $1/u - 1/f = 1/v$.

If $1/f > 1/u$ then the remaining divergence is clearly negative, a fact which signifies "convergence." Similar considerations apply to diverging lenses. This outlook is quite independent of our previous convention of signs. If taken earlier, instead of the treatment previously given, it scarcely brings out the limitations of the thin lens theory and so has not the same educational value.

When will examiners and textbook writers cease grouping lenses as "convex" and "concave" and adopt the more rational "converging" and "diverging" grouping?

With regard to optical instruments only the elementary principles can be given in schools and university first-year courses. In telescopes, for example, the reason for long focus object glass and short focus (high power) eyepiece can be given, correction for chromatic aberration can be dealt with fully and spherical aberration to some extent. A mention that there is no advantage in increasing the power of an eyepiece beyond a certain value and that there is great advantage in increasing the aperture of the objective joins with the remarks arising from the pin-hole camera in preparing the mind for future work.

(4) *Inverse square law*. There is need for drastic modernising of the teaching in this grouping. The teaching must deal with a unit standard source of light and two derived units—the unit of luminous flux (the lumen) and the unit of "illumination" (foot-candle or metre-candle) and the relation between these.

Mean horizontal and mean spherical candle-powers should be introduced.

In photometry we still meet—even in new textbooks—the instruction to adjust so that the "grease spot may *disappear*" on both sides at the same time! Moreover, the inverse square law is stated "for a point source" and then, all too frequently, used in photometry where the sources are by no means points without the slightest reference to errors introduced if the distance from source to photometer is small. The "ten times" rule used in accurate photometry should be known. Many serious students have been worried by this trouble.

Few textbooks mention the above units (i.e. the lumen, etc.) and many teachers may be unfamiliar with them. Those textbooks which do mention them usually give unnecessarily complicated definitions and introduce unnecessary units such as the "lux."

The author would like information as to the best way to present the unit standard source of light. Supposing we say a unit source is a point source of unit candle-power radiating uniformly in all directions; some such definition seems necessary in order to present the "lumen" satisfactorily, but how is it best stated in terms of a fundamental standard? Is the fundamental standard in this country still the pentane lamp? Can the above standard be stated as such that if ten be grouped together they would emit per second per unit solid angle the same quantity of light as is emitted per unit solid angle per second in a forward direction by the pentane lamp?

The "scholarship boy" can be introduced, firstly by experiment (colours of thin films, etc.), to the whole range of optical phenomena which come outside the

above four groupings. That all these facts can be satisfactorily correlated if light be regarded as wave motion gives sufficient incentive for a close study of wave motion, and by this time the student is mathematically equipped to tackle it satisfactorily. A return should then be made to elementary facts to show that the assumption of wave motion is not inconsistent with the "laws" of reflection and refraction. The student should be warned that such a simple wave hypothesis cannot account for many facts concerned with absorption and emission, so that it probably represents only one incomplete aspect of the truth.

11. THE TEACHING OF GEOMETRICAL OPTICS BY MEANS OF THE GAUSS SYSTEM WITH SPECIAL REFERENCE TO THE FORMATION OF IMAGES

By CONRAD BECK

ABSTRACT. The teaching of optics should be based upon the principles of image formation treated according to the Gauss system, the nature of the approximations which are involved being duly explained. The Cartesian convention as to signs should be applied, with the origin at a suitable fixed point in the system while the light is always considered to come from the left. No convention should be used in which the signs change with the nature of the lens.

I CONFINE my remarks on elementary optical education to that branch of the subject which covers lenses and optical instruments composed mainly of lenses. The teaching of physical optics is of course equally important, and it is assumed that the principles of refraction, reflection, interference, diffraction and polarised light have been taught in an elementary manner.

I strongly recommend that the geometric principles of dioptric instruments should be taught by means of rays and not by wave-fronts. I have not yet seen a system dealing with the subject by means of wave-fronts which lends itself to the explanation of the formation of images by lenses and lens systems. Many years ago I studied the ordinary Cambridge text-books of those days—Parkinson—Heath—Hermann, etc., and it was not till I read a German book on the Gauss theory that I ever obtained a mental picture of the action of a lens system. It may be that a system equally simple and equally illuminating could be devised in which the wave-front method is employed, but until that has been done I strongly urge that lenses should be treated with special reference to the formation of images, and that the subject should be treated by the Gauss system. This system is of course based on approximations which must be clearly explained, and it must be understood that the theory is only an imperfect hypothesis. It can, however, be applied without sensible errors to so many optical instruments that it is invaluable for presenting to the student a concise description of the action of an optical instrument.

The following characteristics can be dealt with: Position and sizes of images. Magnifying power. Illumination and photometry. Entrance and exit pupils. Field of view. Aperture. Working distance. Perspective.

I suggest that simple lenses should be considered on three methods with special reference to the formation of an image:

(1) The images of points on the axis of the lens.

(2) The direction of several axes of oblique bundles incident from selected points on an object which are not on the axis.

(3) The consideration of one complete bundle of light from a point on an object some considerable distance from the axis.

Combinations of lenses should then be considered on the same three methods, and these methods should be applied to at least six typical optical instruments to find the optical characteristics of such instruments. Achromatism can be roughly worked out by the Gauss theory, and then, but not till then, aberrations and the general principles of their corrections should be explained. Their correction by calculation, except in the case of achromatism, should not be attempted in any elementary education.

The analytical and geometric work must be done on a reasonable convention as regards signs. Light should always be considered as coming from left to right. Distances to the left of the origin should be minus, to the right, plus. North should be plus and south minus. The origin should be the non-variable point and in general this is the vertex of the refracting surface or the equivalent plane. If the equivalent planes are separated their separation is plus, and if they are crossed it is minus. No system of signs should ever be used in which the convention as to signs change according to the nature of the lens or optical system. Such systems lead to infinite trouble.

For a complete description of this method of reckoning signs, I would refer to Ferraris on *Dioptric Instruments**.

The problems of reflection can be treated by means of dioptric formulae if μ be considered as equal to -1, but the question whether this device would work out conveniently for general purposes would require some consideration on account of the reversal of the direction of the light.

I consider that the Gauss system is of such value in giving rational understanding of lenses that I should propose teaching the general case first and taking the thin lens as a special case where the equivalent planes coincide and teaching both the principal and nodal planes, and taking the lens with air on each side when the combined nodal and principal planes become the equivalent planes as another special case.

There are certain general ideas that require special attention in the above teaching. It is important for the student to understand that rays of light and points are not geometrical lines and points but are only used as a means of finding positions. The limitations of the Gauss theory should be thoroughly recognised, and the conditions under which they can be suitably applied should be understood.

I do not consider that elementary education should be purely experimental as advocated by Mr T. Smith. If no analytical method were available, except one requiring advanced mathematics, there might be something to be said for such a plan, but algebra and a very slight knowledge of trigonometry are sufficient for the treatment of image formation by the eye and most optical instruments; and it is the proper appreciation of the principles of image formation that should form the basis of general optical education. The special work on the correction of aberration and the details of the design of instruments is only required by a few specialists and cannot form a part of elementary education.

* Faber, H.M. Stationery Office.

12. GEOMETRICAL OPTICS AT THE NEWER UNIVERSITIES

By W. EWART WILLIAMS, M.Sc., King's College, London

ABSTRACT. The various grades of geometrical optics teaching at King's College, London, are briefly outlined. Experience has shown the Abbé method, as developed in Drude's *Optics*, to be the most instructive method for both the special and general honours courses. It is suggested that the Physical Society should set up a representative committee to select a system of sign convention and notation and press for its general adoption.

It is with considerable diffidence that I yield to the persuasions of the assiduous Hon. Secretary of the Physical Society to make a contribution on this subject; it is only the realisation that I have spent several years on the industrial as well as the academic side of optics that has finally persuaded me to write these notes.

The following remarks refer in particular to the undergraduate teaching at King's College, London, but are probably applicable in general to the majority of the newer universities. We have to recognise three separate grades or standards of study:

§ 1. SPECIAL HONOURS COURSE

(*A*) Honours B.Sc. (special honours course at the University of London).

(*B*) Pass B.Sc. (general honours, the University of London).

(*C*) Intermediate (including 1st M.B. and Conjoint Board Examination).

If the dovetailing of these separate courses can be improved and a uniform system of notation and sign convention be adopted, this discussion by the Physical Society will not have been in vain.

Considering first of all the geometrical optics teaching in the *A* class, it is an unfortunate fact of which no one will be more conscious than myself, that the student after completing the course is not able to carry out lens computations and would apparently be useless in the lens designing department of an optical factory. It is probably this fact that has given rise to the discussion. Many will have heard the complaint of the universities' failure in this respect. It comes mainly from those manufacturers who expect to have a plentiful, and therefore eventually cheap, supply of optical designers. The old reproach of Lummer and S. P. Thompson that geometrical optics was a mere side line of mathematics is no longer true. We no longer find examination questions that only require a knowledge of Snell's law and a certain amount of ability in algebraic and trigonometrical transformations. The problems of the rainbow are left entirely to the mathematics classes.

The real source of the trouble is that in preparation for an honours degree in physics, only a very limited amount of time can be allocated to geometrical optics. In fairness to the students, the lectures and practical work in the different branches

have to be divided according to the relative importance attached to them as shown by the examination questions. There is as far as I can see no reason for thinking the present division unfair, but as measured from the examination papers of London University, the whole of "light" only represents one-fifth of the two years' (post-intermediate) course. Geometrical optics is represented by about one quarter of this, the remainder being physical optics and including the electronic theory of dispersion and series spectra. *The selection of questions is such that a candidate can obtain full marks without attempting any geometrical optics questions.*

It is obvious that if a student is to be of immediate use in an optical workshop his training must be much more extensive. He could profitably spend the whole of the two years on geometrical optics together with the necessary subsidiary mathematics. On the other hand, if such courses were generally available at our universities I doubt if we should have the students, for the opportunities of suitably placing them afterwards would be very few. The only student that I have had during the last nine years who has shown a particular liking and aptitude for this subject, and who, for pure interest, has subsequently followed it up in considerable detail, is now engaged in an electrical laboratory. I failed to find a suitable opening for him in the lens computing work that he preferred.

The needs of the industrial world must be fully satisfied by the supply from the Applied Optics Department of the Imperial College, where the subject is studied by itself and not as a very small branch of physics.

In the short time available the best course, in my opinion, is to follow the treatment of Abbé's theory as given in Drude's well-known text-book. The student's grasp of the subject should be checked by letting him calculate the equivalent focal lengths and the positions of the cardinal points for actual doublets as well as for the "examinational" hemisphere immersed in oil tanks, etc. Practical work on a photographic objective, with the use of Dr Searle's instructive apparatus, helps considerably in consolidating the theoretical work. No systematic analysis of aberration (other than chromatic) is made, but the importance of the sine law and the effects of apertures are emphasized. In this connection I would like to mention that a few weeks ago I was examining a special instrument, the optical parts of which were made by a well-known British optical firm. In order that it might be used successfully, the illuminating lamp had to be so overrun that it had a life of a few hours only, yet the numerical aperture of the system used was comparatively low.

Finally the adjustments of a spectrometer are carefully considered; this is especially needed because no adequate treatment is given in any of the usual text-books. Practical work with an accurate spectrometer that is put out of adjustment before the student begins to use it is very useful in this connection, provided that only correspondingly accurate results be accepted. The training thus gained is of value throughout the whole of physics.

It would be interesting to study the defects of images in more detail, for example along the lines of Prof. Whittaker's delightful monograph in the Cambridge Mathematical Tracts, but with the time available this is not possible. I believe it is better to concentrate on first order theory as outlined than to aim at a necessarily scrappy

general picture of the whole subject. The student entering an optical workshop has nothing to "unlearn" and his familiarity with first-order work will always stand him in good stead. I would mention here that until some further generalisations have been made in lens theory the real practical work of lens computing will not be materially lessened.

§ 2. GENERAL HONOURS COURSE

These students are expected to understand the essential properties of thick lenses and the significance of cardinal points, etc. During the last few years I have experimented with the various ways of presenting the subject. The most successful method, involving the least amount of pure memory work for the student, has proved to be a simplified form of Abbé's theory. The collinearity equations (in two dimensions) are derived by trial from the easily accepted facts that only one point image of a single point object is required and that the image co-ordinate must be finite (at the focus) when the object is at infinity.

In dealing with nodal points it seems well worth while to demonstrate the reality of their existence and their properties. This is simply done by placing a blackened slide, on which a cross mark has been scratched, in the lantern and adjusting the objective so that the slide is at its focal plane. An aerial camera lens of about 1 metre focal length is placed on a turn-table and the image is formed on a small screen. The absence of any displacement when the lens is rotated about the nodal point (obviously not the centre of the system) is very striking, and can be shown to a large class by substituting a suitable objective for the hand screen and projecting the image on to the wall.

Apart from a descriptive account of the aberrations (chromatic aberration is studied in more detail) this, together with a brief outline of the chief optical instruments, is all that can be covered in the time. An additional advantage of this method, that was not foreseen at the beginning, is that several students who complete their general honours course in two years spend their third year in special honours work. The continuity of treatment thus effected is of great benefit to them not only in consolidating their earlier work, but in giving them more time for further individual study.

§ 3. INTERMEDIATE COURSE

The only comment I wish to make with regard to the intermediate and the corresponding higher school stage is this. For the sake of those students who may be studying the subject further, the sign convention should be altered so that it is in agreement with that used in their later studies. I have seen many examples of the confusion due to the change, and well remember my own difficulties in reading a now classical English treatise on the subject, due merely to the unique convention adopted in it.

The convention that we use is to consider a surface positive if convex towards the object, while distances of object and image are positive if they lie on the lens side of their corresponding focal points, which serve as origins. To keep the usual

intermediate convention would needlessly complicate the higher work while no standard reference book using that notation would be available for consultation by the student.

In conclusion I would venture to make the following constructive suggestions:

(*a*) That the examination papers in optics should be so amended that full marks cannot be obtained without attempting some geometrical optics questions.

(*b*) That the Physical Society, in connection with representatives from other interested bodies, should set up a committee to devise a new, or select an old system of notation and sign convention, and press for its general adoption.

13. A COURSE IN GEOMETRICAL OPTICS

By G. C. STEWARD, M.A., D.Sc., University College of Hull

ABSTRACT. Geometrical optics is at present out of favour as a subject for university teaching. In this paper a suggested course in geometrical optics is given, embracing the elementary non-aberration part of the subject and based upon the Newtonian formula $xx' = -ff'$ and the idea of geometrical collineation; together with a presentation of the more advanced theory, based upon the characteristic function of Hamilton or upon a modification of this function. Suggestions are also made for bridging the present-day gulf between theory and industry.

THE subject of optics, in its geometrical aspect, has been studied from early times, and that this should be so is natural since light plays so important a part in man's experience of the natural world. And since the seventeenth century, perhaps, the subject has been peculiarly English; over a considerable period of time English glass and English lenses and mirrors held a high reputation, and it is only in comparatively recent times that this country has been subject to foreign dominance in optical matters. Moreover the theory of optical systems was investigated here quite early, and a number of results were obtained and mathematical expressions discovered by English writers—results which now commonly bear other names; for example, the Petzval sum and four, at any rate, of the five aberrations of von Seidel were discovered by Airy and Coddington before the time of Petzval and von Seidel. It is natural therefore that the more theoretical side of geometrical optics should have received some considerable attention, and it follows that the subject, in its more elementary parts, has been worked out somewhat extensively. And partly in consequence of this geometrical optics is largely out of fashion in our day in English universities; but also several other reasons operate to this end. The subject appears isolated—it does not seem at first sight to stand in any close relationship to other branches of mathematical or physical enquiry; the elementary theory of systems of lenses and of the bending of rays of light in their passage through various optical media—all this appears an out-of-the-way subject and removed from the high road to any wide scheme of mathematical or physical education. And further it must be admitted freely that the common method of presentation in academic circles is too often not attractive; for one thing perhaps there is not sufficient contact with practical work, with the handling and use of optical apparatus in the laboratory: and, too, not a few teachers are themselves wholly uninterested in the subject, regarding it merely from the point of view of giving to their students the minimum amount for examination purposes. The consequence of all this is that geometrical optics is ejected from the syllabuses of university examinations whenever possible.

In the present discussion it would seem useful to consider two matters and, in the first place, to consider them separately; namely, geometrical optics as an academic subject studied by reason of its own interest and, it may be added, from the point of view of examinations, and, secondly, as an industrial subject—as carried out in the actual design and manufacture of optical systems. And there falls also to be considered the very definite and wide gulf which separates these two at present, so that we may find if possible some means of bridging the gap. As regards industrial methods in geometrical optics—the various ways in use of computing and designing optical systems (and I understand that these are very highly individual)—I cannot claim very close acquaintance with many of them; although I have done my best to keep myself informed of as many of these methods as may be, and of the practical bearing of optical problems.

With respect to the academic teaching of geometrical optics whether elementary or more advanced (although for the most part as far as I am aware it is almost wholly elementary) it must be admitted that the subject as commonly encountered in the text-books is not attractive—in the form of repeated applications of the formula $\pm 1/v \pm 1/u = \pm 1/f$, the signs of these quantities being (to me) to the last degree uncertain and puzzling. And, inasmuch as the student, at present, is unlikely to pursue geometrical optics in its higher branches, the more fundamental principles of the subject do not ever appear to him. The thing which strikes one in contemplating so many presentations of elementary optics (and by elementary optics here is meant a treatment which neglects the theory of aberrations) is the continued necessity for approximation and the (to me) completely baffling sign conventions which are adopted. Approximation of course there must be if one is to neglect aberrations; but it would appear possible to acknowledge this initially and to realise then that one is dealing with a system of purely geometrical collineation—with a unique one-to-one correspondence, that is to say, between the points, lines and planes of two three-dimensional regions which may be labelled, for the sake of importing optics, the object space and the image space respectively. It is not suggested that the first approach to geometrical optics should be after this manner, whether for the mathematical or for the non-mathematical student; but this view of elementary optics is perhaps the one to which the student should be led eventually. The introduction to the subject might well be made through practical work so that the student learns from the handling and use of lenses and lens-systems in a laboratory that he is dealing with a real subject in the natural world; he should begin here so that he may keep this in mind in his subsequent study of the subject. But the time ought eventually to come when he studies the pure theory of geometrical optics, and it would appear to be a gain—for the more mathematical student certainly and I think for others too—if he realises that in the elementary theory he is dealing purely with a geometrical collineation.

In teaching elementary geometrical optics to undergraduates reading for the Tripos at Cambridge, and elsewhere, I have always aimed at an early exposition of the Newtonian formula $xx' = -ff'$, the principal foci of course being the origins of co-ordinates; together with the associated results $x = -f/m$ and $x' = -mf'$, the

notation being as usual. For mathematical students these results are almost intuitive, but for such students, and also for others having less of a mathematical equipment, I have first worked out various simple problems involving thin lenses and mirrors, seeking always, however, to reach the above results, in the first place as particular cases but finally as a general law for all optical systems; emphasizing at the same time the fundamental character of the formulae and showing carefully that they include all cases which can arise in a consideration of the elementary, non-aberration, properties of the symmetrical optical system composed of lenses and mirrors or of any combination of these. This implies the introduction of the idea of cardinal points and cardinal planes—due to Gauss—in the first place gradually and for particular systems, but finally, and for mathematical students perhaps immediately, in the most general manner; so that the student may learn to regard any particular system as but a special case of the general system. From a knowledge of the positions of any two of the three pairs of cardinal points, namely, the principal foci, the unit points and the nodal points, the solution is of course immediate for any non-aberration problem concerning a symmetrical system. Thus if one finds for example the positions of the principal foci, F_1 and F_2, and of the unit points U and U',

then since by definition $F_1U \equiv f$, the first focal length of the system, and $U'F_2 \equiv f'$, the second focal length of the system, one can find the axial point P' conjugate to any given axial point P by means of the result

$$F_1P \,.\, F_2P' = -F_1U \,.\, U'F_2,$$

and the magnification associated with these points from the relations

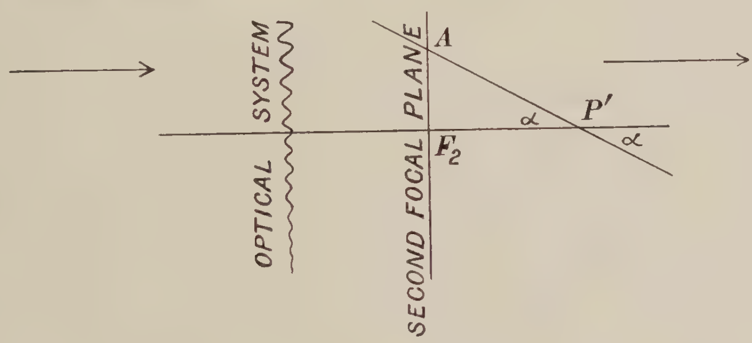
$$F_1P = F_1U/m \quad \text{and} \quad F_2P' = -mU'F_2.$$

For the general optical case this implies an ability to find readily the positions of F_1, F_2, U and U'; and the method is first to write down the power (K) of the system and then the positions of F_1 and F_2 relative to the first and last optical surface respectively, using for this purpose the results already obtained; while finally U and U' may be fixed in position by means of the result $f/f' = F_1U/U'F_2 = \mu/\mu'$, μ and μ' being the optical indices of the end media respectively.

In simple problems involving only two or three refracting surfaces the various formulae can be obtained once for all and quoted as required; but they are so obviously special cases of the general results for any number of surfaces and these general formulae are so compact and easy to remember that it appears to me to be well worth while to give them to the student after a suitable introduction, so as not to leave him only with special cases. And here a word may be said about the use of continued fractions; for the consideration of the general case, involving any number of reflecting or refracting surfaces, continued fractions offer a very

obvious convenience and they can be introduced quite simply even to elementary students. For one thing, only the barest elements are necessary, and I have found that they can be explained *ab initio* as they are required; and the increase in, and feeling for, generality so acquired is I feel of quite definite value. Further, and this applies to students who are to pursue the subject in its more advanced branches, whether academically or from the industrial point of view, the methods lead to a simple way of computing the general system and also its aberration coefficients to be mentioned later. And finally after reaching this general view of the elementary parts of the subject I have commonly taken a large number of examples as problems, whether very simple or more complicated, and have worked them out, or caused the students to work them out, showing that this may for all cases be effected quite simply by the one general method.

A word may be added here perhaps as regards the sign convention adopted; which is that all distances whatsoever are regarded as positive when measured in the direction in which light is travelling through the system. This will imply that angles are regarded as positive when measured as in the following figure:



F_2P' is regarded as positive and also the angle $F_2P'A$.

In conclusion of this part of my subject, and indeed with reference to what follows, I may be permitted to mention a book* of my own recently published, in the first chapter of which is given an outline of the theory of the symmetrical optical system from the non-aberration point of view. The outline is of necessity highly condensed since the book in question attempts to give an outline theory of the symmetrical system both as regards the geometrical aberrations of the several orders, and their computation, and also the effects of these aberrations upon the diffraction patterns associated with the system; and all this in small compass. But I have constantly used the general methods, suitably amplified, for first year students; and these methods have been designed too as an introduction to the higher branches of the subject.

The interest of geometrical optics, however, whether from the academic or from the industrial point of view, lies in the discovery and use of general principles and in the discussion of the actual behaviour of optical systems as contrasted with

* *The Symmetrical Optical System*. Cambridge Math. Tracts, No. 25.

the idealised behaviour of the perfect system presented above; in a word in the aberrations of the system. A general method of investigating these is necessary, and we must also keep in mind the diffraction patterns associated with the optical system and their modifications in the presence of the geometrical aberrations. We are led at once to the wave theory of light, to which, I feel, the student cannot be introduced too early. Not a few of the elementary, non-aberration, properties of the system can be obtained directly from the wave theory, for example, the deduction of the two principal foci and focal lengths and the relation $f \, f' = \mu \, \mu'$*. And it is a gain if the student is taught at an early stage to think of wave surfaces and of the rays of light as being merely normals to these surfaces.

We are familiar with the present-day ideas of quanta in connection with energy propagation and in particular in connection with the propagation of light, but we believe that these quanta have waves associated with them—waves obeying the general laws of wave propagation. The synthesis of these theories is yet to come—or perhaps is now coming through the new wave mechanics; but for a general point of view in geometrical optics we may fall back upon the wave theory. And the student finds intuitively evident the constancy, for a path variation, of the length of the optical path between two accurately conjugate foci, when he visualises waves expanding from the one and contracting to the other after successive reflections or refractions. Further, the mere idea of waves, combined with the conception of interference, leads at once to the idea of a unique geometrical path for a ray of light between two arbitrarily assigned points, through the intervening optical media; that, namely, for which neighbouring paths interfere least, or that for which the number of wave-lengths is stationary for a slight variation of geometrical path. Hence at once the principle of the stationary path between two points P and Q, in the usual notation,

$$\delta \sum_P^Q \mu \delta s = 0 \quad \text{or} \quad \delta \int_P^Q \mu \, ds = 0,$$

where μ is the optical index, proportional to the reciprocal of the wave velocity, for the medium in which the element of path δs is measured; and the symbol δ denotes an arbitrary, small, variation of the geometrical path between the end points P and Q, which themselves remain fixed.

I do not think that any elaborate apparatus or mathematical knowledge is necessary to give the student a clear understanding of this principle which is purely physical in origin and character; and the simple elementary results for an optical system may be obtained readily from an application of it while of course the general theory may be obtained from it immediately.

Sir William Hamilton embodied this principle in his characteristic function and in the several modifications which he gave of this function, including that one now known as the Eikonal and associated with the name of a subsequent investigator, Bruns. It was pointed out by the late Lord Rayleigh how readily an examination may be made of the first order aberrations of the symmetrical optical system (and equally one may add of the aberrations of any order) by means of this

* Cf. *The Symmetrical Optical System*, Ch. II, § 3.

haracteristic function; and other writers too have dealt with this point of view, for xample, Clerk Maxwell, Larmor, Bromwich, T. Smith, etc. It is not my province o enter here into details of the application of the characteristic function or of the ikonal; I may perhaps refer again to the book mentioned before where a connected account of these matters is given. It will be sufficient simply to say that by n expansion of either of these functions in terms of suitably chosen variables (and great deal lies in a suitable choice of variable) an examination of the geometrical berrations may be made quite readily, and the nature of each of them investigated, or aberrations of any and all orders. The general method is to evaluate, for example, he eikonal for a hypothetical system, supposed perfect, that is, free from aberrations for two conjugate planes at a given magnification m, and then to evaluate it for the general optical system; the difference between these two results, named the aberration function and introduced by T. Smith, sums up in itself all the aberrations of the system and depends upon these aberrations alone. We have here a compact, ready and convenient method of investigating the qualitative and also the quantitative nature of the geometrical aberrations of all orders; from the practical and industrial point of view it is necessary of course to consider them quantitatively, and this too may be effected by means of the aberration function; it is here perhaps that we may bridge the present-day gap between theory and practice. We find formulae giving the aberrations of a single surface and then further addition formulae to give the effect of the addition of several surfaces; in this way obtaining the total aberration effect of a combined system in terms of the separate effects of the several components of the system.

And finally there remain the problems concerning the diffraction patterns associated with optical systems and their modification in the presence of the geometrical aberrations; these again may be investigated very readily by means of the characteristic function. A more elaborate mathematical equipment is necessary here if the problem is to be examined in its generality; although some progress may be made through numerical quadratures of the integral expressions obtained. But for an outline of this I may refer again to the book already mentioned and to the references given there.

The preceding gives a very brief outline of the method which I have adopted in teaching students the more advanced parts of the subject of geometrical optics; and I think that I can say that the students have been interested in the presentation.

14. OPTICS IN THE INDUSTRY

By JAMES WEIR FRENCH, D.Sc., F.Inst.P.

ABSTRACT. The subject is discussed only from the industrial point of view. The first point raised is that educational authorities cannot subordinate more important interests, such as those of the chemical and electrical industries, to the minor claims of the optical industry which, although important, is comparatively small. A statement of the needs of the industry leads to the second point, that, as the work of an optical designer so largely depends upon his workshop and general experience, his training must come from the industry itself, and that only a good, comprehensive education including a knowledge of fundamental optical principles is necessary as a preliminary to workshop training.

I take it that the purpose of this discussion is to outline a course of optical study that will be of general educational value, that will be of service to the industry, and that will help the user of instruments to appreciate their optical performance.

From the industrial point of view, I should prefer to consider first what type of education best suits our requirements. Before doing so, I cannot refrain from suggesting that difficulty must be experienced, not so much in the formulation of a course of study as in the reformation of what exists, since such reformation must involve an increase in the number of lecture hours.

Here and there throughout the country some individual teaching of optics may be found, but in general it forms no important part of a school curriculum. If we exclude the admirable practical teaching of Prof. Conrady at the Imperial College and the ophthalmic teaching at the Northampton Institute, the subject of optics—other than physical optics suitable for examination purposes—forms a very small part of the teaching of our colleges and universities. Our physicists and mathematicians are not disposed to devote their time to the material aspects of optics. All who can do so are more profitably employed groping among the ruins of the 19th-century edifice of science, searching for the architect's plan and building new foundations here and there where the ground seems reasonably sound, hoping to find a ray of light among the débris.

Educational authorities would no doubt be very willing to teach everything well, if circumstances permitted. But they are unable to do so. From all sides they are pressed to allocate more hours of study to cope with the rapid development of chemical engineering and the electrical industry. Three-year courses have been extended to four and hitherto essential subjects have become alternatives. Reformation of optical teaching means an extension of hours and, when one considers the strength of other appeals, it is difficult to know how these can be allocated, either in schools or colleges.

My first point is that educational authorities cannot subordinate more important interests to the minor claims of the optical industry which, although a key industry, is a comparatively small one.

It may help us to form some idea of the kind of optical instruction which is desirable if I outline the work involved in the design of an optical instrument. In small concerns such design may depend upon the enthusiasm of an individual of

outstanding ability and ingenuity. To an ordinary mind computation work can be appallingly monotonous and there are many men who, having the other qualifications, fail in this particular respect. Such an individualistic system is only of limited application. In larger establishments the individual must be supported by a number of assistants. I recollect that, in the case of one important continental firm, the material prepared by one individual was handed over to about thirty or forty computers, expert in the use of logarithmic tables but otherwise of no very special education. This was before the introduction of computing machines which have reduced so materially the actual labour of calculation. The design of an optical instrument is not the work of such computers. The actual designer may not be skilled in this particular section of the work, for the reason that the mechanical may dominate the optical. When the optical system as a whole has been designed, the computation of the elements, individually and in combination, is entrusted to the operators of the calculating machines who need have no exceptional mathematical equipment other than a common knowledge of the simple trigonometrical expressions, which can hardly be regarded as mathematics. Particular rays are traced step by step through surface after surface for the purpose of determining at various stages the longitudinal and transverse aberrations. These values are assessed by the skilled computer who decides at what particular part of the system a modification can best be effected. His special skill is much more practical than mathematical. It is based essentially upon industrial and commercial experience that can be gained only from years of active practice within the industry itself. It depends upon a knowledge of what is necessary to excel the productions of progressive competitors. It involves experience of workshop conditions and, above all, good judgment in the balancing of one type of aberration against another, for no optical system can be free from all kinds of aberration. To satisfy the needs of some particular client, it may even be necessary to re-assess the relative importances of the various aberrations to suit the ideas of the customer.

Some particulars regarding a typical submarine periscope of average dimensions will afford an indication of the magnitude of the work involved in the computation of the optical system. Of simple operations involving, say, two multiplications and one division, there are 12,000 in this particular example, which is not an extreme case; the number of individual operations is about 40,000; and the amount of writing involved is equivalent to the writing of a 70,000 word book. When it is considered that the instrument in question has to be designed and constructed within the limits of a contract date with the probability of a penalty for delay in delivery, it will be realised that in actual practice such work cannot be undertaken by an individual; it must be so organised that it can be distributed over a group of computers.

So far as the general design is concerned, methods no doubt vary greatly. My own particular system is somewhat as follows. The most suitable way of meeting the specification having been decided upon, an optical diagram showing the positions of the lenses, the central beam, and the oblique beam, is prepared; this path-in-air diagram is adjusted to suit the principal rays determining the angular field and the cones of rays determining the illuminations at the centre of the field

and at the margin. Upon this diagram the thicknesses of glass are then indicated and the diagram is adjusted accordingly, the amount of adjustment generally being small as, with experience, allowance can usually be made in the initial diagram for the air/glass differences. The principal thicknesses having been determined, the system is redrawn with greater precision, the final dimensions being checked by calculations, all of the simplest arithmetical kind. The designer knows from experience the difficulty that may be incurred in the computation of each particular element and his principal aim is to ensure that he does not set the computers too difficult a task in any one respect. There is great scope for ingenuity in the preparation of this original design and it is in this particular stage that the most effective work is done. Particulars of the system are then sent to the computer. He receives information regarding the focal lengths and diameters of the various elements, approximate thicknesses of the lenses and prisms, and air separations. He also receives for the axial beam its radius at the entrance surface of each element and for the extreme oblique beam the radial distances of the points of intersection of the principal ray with the respective surfaces. The approximate radii of the individual surfaces of each combination are then calculated by simple proportion from some existing similar combination; it is rarely necessary to compute the curves *de novo*, but even if that were necessary the ordinary simple approximate formulae would suffice for the purpose. The approximate curvatures having been determined, the work of computation by the now well-known trigonometrical system can then be commenced.

Practical considerations then determine the progress of the work. Theoretically it may seem desirable to modify the surfaces of one of the large diameter objectives or projectors. This, however, would involve the construction of large and expensive test plates. Practical considerations make it advisable to manipulate the smaller diameter entrance or exit systems.

My second point, therefore, is that the work of an optical designer so largely depends upon his workshop and general experience that his training must come from the industry itself and that only a good comprehensive education, including a knowledge of fundamental optical principles, is necessary as a preliminary to his workshop training. A schoolboy's education must of necessity be general, as the kind of work he will be engaged upon later is quite unknown. The more determined he is to become an optician, the likelier he is to become something else. Of the hundreds of boys of whom I have had experience, I cannot remember one who possessed more fundamental knowledge of optics than could be acquired by him in a few evenings of purposeful study. And so far as the kind of optical knowledge required by the industry is concerned, the same applies to graduates. I personally have long since ceased to expect any such specialised knowledge. I prefer to remember that inherent qualities may be of more permanent value than those acquired, that the steel industry was revolutionised by a clerk, that after the Fire of London an optician was entrusted with the design of St Paul's, that James Watt—another optician and instrument-maker—became the greatest of steam engineers, and that on the other hand we owe the art and practice of optical glass making to a carver of clock-cases.

15. THE NEED FOR MEETING PRACTICAL REQUIREMENTS

By Instructor-Captain T. Y. BAKER, F.Inst.P., R.N.

ABSTRACT. The subject should be taught so as to be of use both to users of optical instruments and also to designers. The course should be mainly Gaussian optics with only a qualitative exposition of aberrations. Much more design work of a Gaussian character ought to be done, particularly with the view of ensuring requisite diameters to lenses. Examples of optical systems should include more modern instruments and not be virtually confined to the astronomical telescope. There is scope for a far greater development of the analysis and design of reflecting prism systems.

Optics, like many other branches of applied physics, leads to a definite technical end, the designing of optical instruments. In teaching the subject, sight must never be lost of the needs of the professional designer. At the same time—as in other subjects—optics is of deep interest to the many who have to use optical instruments. In numbers they exceed the technical class, and it is obvious that the general education in the subject does not need to go so far for the user as for the designer. Up to a certain point the course for both classes can be identical and I propose to discuss the methods of teaching geometrical optics up to that point and no farther.

In most branches of physical science the technical development of the subject is based upon certain fundamental laws which are incorporated into some sort of mathematical treatment, using the word mathematical in a very broad sense. In nearly all cases it is the teaching of the mathematical treatment that forms the bulk of technical education. The simple laws do not carry one very far technically. It is the practical methods of making use of them and computing their effects that the technical man requires. For example, the general nature of the magnetic field associated with a current flowing in a wire is easily understood by the young student, but he has a long way to go before he can use his knowledge in the design of a three-phase motor.

Now in geometrical optics there is nothing more in the way of physical laws than the two dealing with the reflection and refraction of light and the mathematical treatment involves, at the most, three-dimensional geometry. Except for plane reflections, the whole development of the subject reduces, as far as the technical designer is concerned, to the theory of the symmetrical instrument. Gauss showed that for paraxial rays the relationships between object and image points are contained in the equations

$$Axx' + Bx' + Cx + D = 0 \qquad \ldots\ldots(1),$$

$$y' = y\,(Ax' + C) = y/(Ax + B) \qquad \ldots\ldots(2)$$

with the relationship $BC - AD = 1$.

In these equations x, y are the co-ordinates of a point in the object space and x', y' of the corresponding point in the image space. The equations form a geometrical collineation, and transform points into points and straight lines into straight lines. They are not strictly true and are approximately so only as long as the fields and apertures that are dealt with are small. But, all the same, they represent what the designer aims at even for large fields and large apertures and his chief concern in computing a system is to choose the lens curves, separations and glasses so that the image defects are kept within tolerable bounds. The preliminary design of most optical systems is, however, made on the supposition that the Gaussian relationships hold for the full aperture and the full field.

There are, of course, many ways in which these relationships may be written. For example, with a suitable choice of origins the first equation is the same as the familiar $uv = f^2$, but no matter what origins of reference be chosen the Gaussian equations form the basis of optical design for all sizes of field and of aperture.

Hence it seems to me that in teaching the subject of geometrical optics more design work on Gaussian lines might be attempted and I should like to see more attention paid to the lateral side of the design and not so much on the longitudinal. I can best illustrate what I mean by a simple example of the type of design work that my own department is continually doing. I need, say, a telescope of the inverting type which has to be used in rather poor light. The telescope is to be used for some purpose of measurement and a knowledge of the sensitivity required suggests that the magnification should be about 10 ×, while the light conditions indicate that an exit pupil of not less than 5 mm. should be obtained. We arrive therefore at a 50 mm. objective, and a knowledge of what can be done in the way of correcting the aberrations of the objective and the eyepiece indicates that the objective cannot comfortably be shortened beyond $f/6$ or in other words that its focal length must be at least 300 mm. This requires that the equivalent focal length of the eyepiece shall be 30 mm. and, knowing that the apparent field cannot be greater than from 50 to 55°, one can decide immediately that the diameter of the field upon the focal plane will be 30 mm. One must then decide what type of eyepiece, Ramsden, Huyghens or Kellner, is going to be used and the Gaussian design of the whole optical system can be completed, drawing it to scale and determining the necessary diameters of the lenses for transmission of the full beam of light.

It is lateral design of this character that forms a useful introduction to the technical end of geometrical optics. I do not assert that this side is never considered at present, but when looking through some of the text books I am constantly struck by the preponderance of examples on the longitudinal side, such as complicated formulae for the apparent distance. I always have the feeling that the authors are using the Gaussian image formation as a convenient peg upon which to hang a lot of complicated algebra. If writers and teachers want to exploit geometrical optics as a subject for the elaboration of mathematical conundrums then, in my opinion, it is better to go further into the only part that gets away from the symmetrical optical instrument, namely the part connected with plane reflections. The geometry here can be quite complicated and to treat it fully a more extensive know-

ledge of spherical trigonometry is required than is usually possessed by the technical optics designer. If the mathematician wants a new scope for his art, that can lead to practical results, I would draw his attention to Mr T. Smith's recent work on *The Theory of Matrices* as applied to the analysis and synthesis of reflecting prism systems.

Far more might be done in the way of instruction in the use of prisms in instruments without any loss to the technical man. Reflecting prisms enter into so many optical instruments and there is any amount of scope for a text book which treats them in a thorough and systematic manner. Such things as the fundamental necessity for an even number of reflections if a right-handed image is required and an odd number for a left-handed image, or the development of prisms into parallel blocks of glass, or the necessity of at least four reflections to cope with the displacement of an object to *any* image position: all these points are of vital interest and are all based upon fundamental conceptions of geometry that ought to be adequately treated. This branch of the subject is full of useful practical examples that can be based upon the more commonly used forms of reflectors, such as the Porro prisms for binoculars, the Prandl prism for rangefinders, the tetragonal or corner-cube prism for returning a beam parallel to itself, and so forth, all of them examples of common practice which every student, whether potential user or potential designer, ought to be thoroughly familiar with.

Again, I wish to urge the desirability of bringing the descriptions of optical instruments more up to date. The telephoto lens in photography, the internal focussing telescopes of theodolites, the rangefinder, the prism binocular, the prismatic astrolabe of the geodetic surveyor, the level, the tacheometer, the kinema projector, the process camera, the variable power telescope: here are a number of optical systems in common use embodying a great variety of Gaussian designs that might, with profit to the user and to the designer, be included in a course on geometrical optics. There is scope here for very valuable instruction in the outline design of instruments, particularly of the lateral character that I have previously mentioned. As a side line in instrument design there is the very important question in many visual instruments of the character of the field of view. Many visual instruments require observation to be made by bringing two optical images into contact. The fields may be superposed as in the case of the sextant, or adjacent as in the rangefinder. Optical systems that will give fields of these two types are of importance, and the current practice for the one type or the other might very well be taught. Again, in instruments used for accurate measurement it is very frequently the case that one wants means of measuring small angles by movement of optical parts with open scales. The various methods of achieving this end, the travelling wedge, the sliding positive and negative lenses, the parallel plate micrometer, the Zeiss rotating wedges, are all subjects that give an opportunity for theoretical treatment that is of great practical value.

I have said enough, I think, to indicate my opinion that the teaching of geometrical optics might be brought much more into contact with reality by including a large number of modern instruments and the gadgets that are associated with

them. I am confident that if this is done the course could be made more interesting to the mere user and the technical man could get earlier into touch with the practical side of his subject. It is possible the inclusion of even a part of what I consider desirable may necessitate the excision of something that is already taught. If so, then I should, without question, omit entirely any attempt to teach aberrations except qualitatively. The instruction must be given in such a way that the student does not lose interest. No doubt every lecturer on the subject has his own ideas, but I think it can do no harm if I put down here the methods that I have used and found to be satisfactory.

Quite early in the course, in the investigation of the Gaussian equations for image formation, the student appreciates the fact that in each successive space between refracting surfaces there is, somewhere or other, a field plane FF' and an aperture plane AA' (Fig. 1), and that in the symmetrical optical instrument one is interested only in a circular part of the field plane and a circular part of the aperture plane, both circles being concentric with the optical axis. Gaussian formulae are only approximately true; they require that the radii both of the field and the aperture should be small. If we take any ray entering the instrument, it can be shown that in the space we are dealing with the ray should travel from P in the aperture plane to Q in the image plane, crossing these planes at heights x and y from the axis.

Since the Gaussian equations are not exactly true, the ray will depart from its ideal position PQ to $P'Q'$. It has a lateral aberration $\underset{\sim}{QQ'}$ on the image plane. It is clear that $\underset{\sim}{QQ'}$ must be a function of both x and y. It is clear also that by rotating the instrument through 180° round its axis that $\underset{\sim}{QQ'}$ must change sign without alteration of magnitude when the signs of x and y are changed without alteration of magnitude; in other words, that $\underset{\sim}{QQ'}$ is an odd-valued function of x and y mathematically small in comparison with x and y. We can therefore put down immediately

$$\underset{\sim}{QQ'} = Ax^3 + Bx^2y + Cxy^2 + Dy^3 + \text{fifth and higher powers.}$$

Omitting for the moment the fifth powers and dealing with what are called first order aberrations, one gets four terms Ax^3, Bx^2y, Cxy^2, Dy^3 which can be utilised to illustrate the nature of aberrations. If the higher order aberrations can be neglected then in the final image space the object of the designer is to make the coefficients zero.

If we take the first term, Ax^3, alone we can see the effect on the formation at the middle of the field. Take two rays, one through the top and the other through the bottom of the stop. They intersect on the axis, but at f', whereas a similar pair of paraxial rays intersect at f. Calculations of the position of f' lead easily to the determination of the simple caustic and the nature of spherical aberration becomes apparent. I find it convenient at this stage to call attention to the shape of the wave-front as it emerges from the stop. In the absence of spherical aberration the shape is that of a sphere. With spherical aberration as shown in Fig. 1 the surface has the same radius of curvature on the axis but flattens out as one leaves the axis.

After spherical aberration it is, I think, more convenient to pass directly to the

astigmatism term, Cxy^2. The trigonometry shows that if the two rays through the top and bottom of the stop be drawn (omitting all the other three terms in the aberration) they intersect at a point Q_1 (Fig. 2) and that as Q is moved up and down

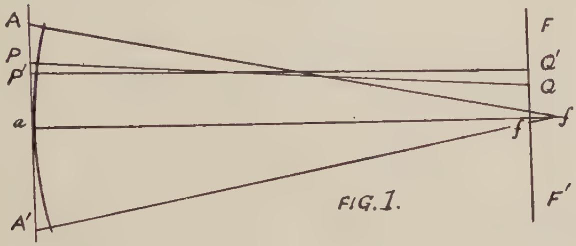


FIG. 1.

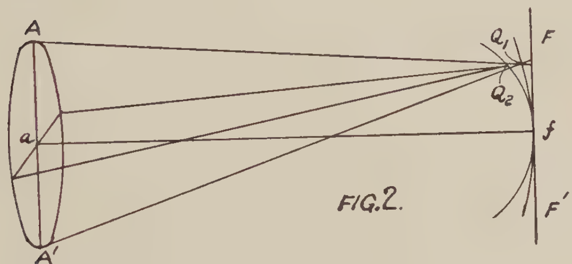


FIG. 2.

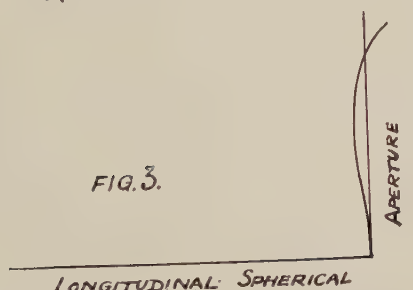


FIG. 3.

FIG. 4.

the field the focussing point traces out a curve of radius $1/C.af$ and we get immediately the notion of curvature of field. The next point is more difficult and requires a three-dimensional consideration. The wave-front emerging from the aperture

when the image point is on the axis is symmetrical with respect to the axis, although it is not spherical. The wave-front converging to a point off the axis cannot be assumed to be symmetrical with respect to its principal ray but the symmetry of the instrument shows that the principal planes of the wave-front lie, one in the plane of the paper and the other at right angles to it. If, therefore, we were to draw two other rays through the sides of the aperture, they will also intersect one another but at a point $\underset{\sim}{Q}_2$. The student has, or ought to have, sufficient knowledge of solid geometry to understand what is meant by the two principal radii of curvature and by focal lines, and the extension of these ideas to the second field curvature is not difficult. I am afraid, however, that I have never succeeded in finding an easy method of showing that the curvatures of $f\underset{\sim}{Q}_1$ and $f\underset{\sim}{Q}_2$ are invariably of the form $3X + Y$ and $X + Y$, where X represents the astigmatism and Y the Petzval curvature. Usually I have stated plainly that these facts emerged clearly from a quantitative examination of the aberrations and could not be easily demonstrated in a simple manner.

I have said enough, I think, to indicate that, as far as spherical aberration and astigmatism are concerned, they can be explained to elementary students in a qualitative manner. Coma and distortion follow in the same way and the student can thus be familiarised with the five von Seidel aberrations so as to be able to understand the problems that face the optical designer in his attempts to produce satisfactory optical instruments. One cannot go much farther. Practical design involves consideration of higher order aberrations, generally by some system of trigonometrical ray tracing. It involves not the removal of the first order terms but a balancing of their residual values against those of higher orders. But I used to find that in respect of spherical aberration it was worth while carrying the matter a stage farther. The customary method of plotting spherical aberration as derived from ray tracing is shown in Fig. 3. As the aperture is increased in a well-corrected system it is common to find that the under correction of spherical aberration increases up to a maximum and then diminishes. For the extreme rays it is frequently over corrected. Let the student plot even roughly the caustic that corresponds to a spherical aberration distribution of this sort and he has no difficulty in arriving at the three-cusped curve shown in Fig. 4. The emerging wave-front has minimum curvature on the axis and maximum curvature at a zone off the axis. This side-light on image formation is of considerable practical interest because the student is able to appreciate that the circle of least confusion is by no means at the position assigned to it when aberrations of the first order only are dealt with.

One cannot go much farther with aberrations. Any attempt to establish their quantitative values by the use of some form of characteristic function should be kept entirely for the technical student in his advanced course.

GENERAL DISCUSSION

Mr T. B. Vinycomb: I should like to emphasise Dr Searle's remark that the pupil should have precedence in the choice of the method of presentation of a subject. Last summer I marked 250 scripts written at the Intermediate and Higher Schools Examination. In the paper was a simple question on a lens. 247 candidates attempted it and 7 solved it. All the 240 who failed in their attempts used the formula $1/u - 1/v = 1/f$. All four who used $1/u + 1/v = 1/f$ solved the question. As the candidates came from all parts and had different teachers it is clear that the convention of signs used in the first formula is not grasped by pupils at this stage. Those who are going further in their studies can be expected to be able to deal with any reasonable convention of signs, but elementary students should not have their difficulties increased unnecessarily.

Mr W. Bennett: It is probable that few teachers of experience have not at some time experimented in the introduction of beginners to the subject through the method of curvature. It is also probable that very few have retained it as their principal method of approach. The reasons for retaining the ray method have been well put by several speakers, but I should like to point out another advantage. In approaching the simple lens by the ray method we have to indicate, by the approximations that we make, within what narrow limits the simple lens can be expected to form good images, and the pupil is in a position to appreciate the wonder of modern instruments. If we approach by way of the wave-front we can assume at this stage only that the emergent wave-front is spherical, and we have no indication of the degree of exactness of this assumption. Light tracks may be used without any special apparatus. A modern gas-filled lamp, placed with its axis horizontal some way behind a slit fixed to the edge of a drawing board gives a track with which much very useful work may be done. I should like to plead strongly for the retention of the formula $1/u + 1/v = 1/f$. It is much the simplest form to use with the optical bench, and indeed, as has been pointed out, any other form leads with elementary students to endless confusion. The objection is raised that with this convention the formula becomes inconvenient when a ray has to be traced through several lenses or surfaces, but surely the formula is the very worst to use for such a purpose, and in serious computation it is not used at all.

Mr A. G. Tarrant: As I have tried out a number of different methods of approach to this subject during the last nine years, my personal experience may be of interest, since, after all, the proof of the pudding is the eating. At first I tried beginning with the physical nature of light. But later experience has shown me the wisdom of being much less ambitious, and now I begin with the observed facts of light propagation, reflection, and refraction. This has the advantage of giving the students, from the first, ideas that are clear-cut and certain. Experience has shown that my students, at any rate, follow this order very much better. To begin in this

way with incontrovertible results of experiment has the added advantage of emphasising a very important point. For it is the function of theory to explain the results of experiment, and the function of experiment is to find out what are the facts concerning any phenomena, not to illustrate theory. In every science accurate knowledge of fact must precede theory. In these days of rapid advance in theory we are in danger, sometimes, of forgetting this, and we ought to be specially careful in training the rising generation of physicists to keep observation and theory in their proper order and relation.

It would seem particularly desirable in this subject to preserve this order, as if we begin with the physical nature of light we are at once faced with our own ignorance. What is the physical nature of light? Which theory, Fresnel's, Maxwell's, Planck's, Schrödinger's, shall we teach? Matters that are still the subject of theoretical argument would seem best avoided with beginners, and for that reason, if for no other, it has seemed best to me, after mature consideration, to keep to the old-fashioned, but clear, safe and certain method of taking the observed facts first.

COMMUNICATED REMARKS

Dr Terry Thomas: My impression is that it is much more difficult for young boys to grasp the implications of the wave theory. They are much happier with lines and points, and, of course, their geometry can be made useful in this connection. Nevertheless I feel that the teaching on optical instruments is generally not satisfactory and I think the text-books are very largely to blame. I very much doubt if many boys leaving school, or even university, students have a very clear idea of optical instruments. Probably not enough attention is given to the fitting up of such instruments in the laboratories. Too often, if they are used at all, they are used in the complete manufactured state.

Mr F. Record: In the Technical College, Derby, we have used the wave method for teaching the properties of mirrors and lenses for some years and are satisfied that it provides a more complete physical account of the action of such appliances than does the ray method. As the result of our experience, we suggest that the dioptre be defined as a unit of curvature. After having shown that the curvature of a wave-front or surface is measured by the reciprocal of the radius, we seek for a unit of curvature. If we choose for the unit the curvature of a surface whose radius is one metre and call it one dioptre, we find naturally that we have a unit for measuring the power of a lens or mirror. This is so because the power is the curvature impressed by the appliance on the incident wave-front. I would like to say how useful the graphical method described by J. H. Dowell* has proved in teaching the principles of the direct vision spectroscope. With a little care in drawing the surfaces in the sketch of the prism and in the scale diagram in the same direction, the whole of the facts with respect to the direct vision prism are presented in a manner which appeals to students. There seems to be no other method which gives so much information in so compact a form.

* *Proc. Optical Convention*, **2**, 965 (1926).

Mr A. W. Barton: There is one small point which I should like to raise although it is rather off the trend of the main line of the discussion. The subject of photometry has rather gone to seed in recent years as far as school work is concerned, but within the last ten years a very striking application of the law of inverse squares as used in photometry has been made. I refer to the determination of the distance of star clusters and spiral nebulae by the measurement of the apparent brightness of Cepheid variables contained therein. The periodic time of the variable star fixes its illuminating power (i.e. candle power) and so its distance can be calculated. It is in this way that Hubble has estimated the distance of the remotest spiral nebulae at 140 million light years. I feel that such a wonderful application of the fundamental principles of photometry would do much to put fresh life into the subject and should certainly be included in any elementary treatment of the subject.

Prof. L. C. Martin: Why teach geometrical optics at all? There are two main justifications for the teaching of any subject: firstly, its possible educational value; secondly, the utility of the knowledge acquired.

I shall venture to express no opinion on the educational value of geometrical optics considered merely as a branch of higher applied mathematics, but it is generally conceded that the utility of much that has been taught in this connection is extremely meagre. It seems to be the case at present that the mathematical purist is somewhat out of his depth when the more complex practical problems of applied optics are encountered, problems which yield only to numerical methods of the "try, try, again" type, as intelligently conducted as possible. This is not said in order to detract from, but rather to emphasise the necessity of, further mathematical research by those qualified to undertake it. The ordinary science student encounters geometrical optics as a branch of physics; he does not reach the stage of mathematical refinement, but has to be content with elementary principles. It is usually understood that the object of the instruction is the attainment of some knowledge of the principles of the optical instruments which he is bound to employ in observation and measurement.

It is unfortunately the case that the Gaussian lens theory to which the elementary student is early introduced rests on a basis of approximation which is not exactly understood, in the usual way, either by the teachers or the taught. The simple trigonometrical tracing of a few rays through a lens might let in a flood of light on the whole subject. I would never introduce a student to the theory of a lens till he has understood the trigonometrical method of exact ray tracing, till he has explored the conditions of refraction at a single spherical surface, and grasped the idea of the curvatures of corresponding object and image fields in that simple case.

The idea of aberration, a terrifying subject to many, can be introduced in one of the earliest pin-method experiments in ray tracing through a plane surface. Then ray tracing, by the aid of Snell's construction, on the drawing board will give a sense of power and exactness which are entirely lacking in the ordinary approach to lens theory.

Then again, a close connection between physical and geometrical theory needs to be maintained throughout the instruction. The idea of the "optical path"; the elementary image as the meeting point of disturbances arriving in the same phase; the aspect of aberration as optical path difference; all these can be introduced in early stages so that the subject develops in a natural and harmonious manner. The Gaussian theory rests very naturally on a basis such as is developed above, and its limitations can be kept clearly in view; the Abbe equations for magnification and for conjugate distances reckoned from the focal points ought to be among the earliest taught instead of being almost entirely disregarded.

Many of these ideas have been worked out in the courses which are being given in the Technical Optics Department of the Imperial College. Prof. Conrady's work is of such a nature that it must in due time have a far-reaching influence on the subject. His methods are set forth in part in his recent book on *Optical Designing and Computing*, Part I. The present writer is also engaged in the preparation of a manuscript on applied optics in which some of the ideas expressed above are worked out on elementary lines in close harmony with Prof. Conrady's methods, the notation and sign conventions therein adopted being mainly used.

Through the courtesy of Mr Emsley, a general agreement as regards notation has been reached between the Imperial College and the Optical Department of the Northampton Polytechnic. A notation for general use, however, must not conflict with that adopted for the theory of ophthalmic lenses in which the use of capitals is generally to indicate the reciprocal of the length denoted by the corresponding small letter. As I strongly deprecate the use (for ordinary students' work) of complicated notations involving types of different thickness, etc. which may cause confusion, it has been agreed to use capital letters of the cursive form to denote the reciprocals of length, thus

$$\mathscr{L} = 1/l,$$

while L in ordinary type can still stand for a marginal distance in the notation of the computer. The great desirability for a notation is easy writing, a matter seemingly overlooked by many who are fond of introducing complicated types.

Optics has the reputation of being a dry and forbidding subject; in recent years the unnatural separation of the Physical and Optical Societies has not encouraged interest in optics amongst teachers of physics. It is more than time that this separation should be ended by union into two groups of one society with joint meetings of both groups frequently arranged, and the option for members of subscribing to one or both sets of transactions.

SUMMARY OF AGREED NOTATION FOR USE IN ELEMENTARY CLASSES OF THE TECHNICAL OPTICS DEPARTMENTS AT THE IMPERIAL COLLEGE AND NORTHAMPTON POLYTECHNIC

Designation of Points.

Object and image	B, B'	Nodal points	N, N'
Focal points	F, F'	Surface vertices	A_1, A_2, etc.
Principal points	P, P'		

Distances.	Paraxial	"Marginal" values (computing notation)
Object and image from P and P' ...	l, l'	L, L'
Size of object and image	h, h'	H, H'
Incident height on surface	y	Y
Distance from F and F' to object and image	x, x'	X_f, X_f'
Separation of systems	d_1, d_2, etc.	
Separation of adjacent focal points ...	g	
Focal lengths, i.e. PF, $P'F'$ (opposite in sign numerically)	f, f'	
"Vertex focal lengths"	f_v, f_v'	
Distance to tangential and sagittal foci	s, t and s', t'	
Radii of pupils	p, p'	
Radius of curvature	r	
Angles.		
Incidence and refraction	i, i'	I, I'
Ray with axis	α, α'	U, U'*
Refractive indices	n, n'	N, N'*
Magnification $= h'/h$	m'	M'

Reciprocals. Focal powers, curvature, and reciprocals of distances to be expressed in capital letters of the cursive form, thus:

r = radius of curvature.	$\mathscr{R}$ = curvature (dioptres) $= 1/r$.
f' = focal length.	$\mathscr{F}$ = focal power $= 1/f'$.
l = object distance.	$\mathscr{L}$ = vergence to object $= 1/l$.

This notation is not used in the computing classes where $1/r = c$, but although the employment of capitals to denote reciprocals is securely founded in the theory of ophthalmic lenses, it is suggested that the consistent use of the cursive form will save confusion with an equally well established computing practice of using the Roman capital to distinguish a "marginal" distance.

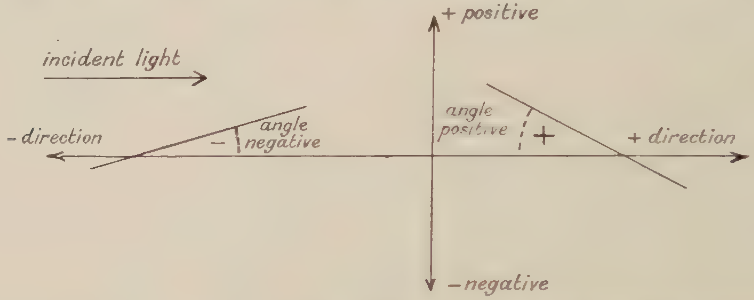
Sign conventions. These are expressed diagrammatically in the accompanying figure.

Surfaces and elements. Surfaces are numbered consecutively 1, 2, 3, etc.; lenses are given letters a, b, c, etc.

* Note that in the computing formulae, the paraxial and marginal angles are u, U. The use of u is to be avoided, however, for elementary students who have been used to the Coddington notation of u and v for object and image distances. Also N is to be used for refractive index in computing classes, and n in elementary classes

Reduced distances. Denoted by adding a bar over the corresponding letter, thus $\bar{l} = l/n$.

It will thus be seen that the notation adopted in the various technical classes will in future have a great measure of consistency; the great aim is to adopt a notation and especially a consistent sign convention so that elementary students may be saved confusion on proceeding to more advanced work and reading.



Mr William R. Bower: A hasty glance through the advanced proofs shows firstly a very notable conflict of opinion, and therefore the urgency of a standardisation of methods, nomenclature, notation, etc. Secondly, that the subject has been handled too much by the non-experimental mathematician, and too little by the physicist, and is therefore terribly susceptible to the hectic influence of examinations. It is certainly time that representative physicists should do something for the betterment of optical education. And may I say incidentally, that not only is optics in distress educationally, but other major branches, especially mechanics, have been over-dosed by mathematicians and the indigestion too easily tolerated by the engineering and experimental physicists. In regard to the teaching of elementary optics, the advanced proofs show that there is, in the main, a conflict between the ray method, with formulated reciprocals, and the wave or curvature method. In my experience the ray method has been the better one. But the wave method when used sparingly is interesting and instructive. The ray method is more easily tackled and when rays have been drawn graphically the wave-fronts are easily traced as arcs perpendicular to the rays. To me the Gauss generalisation of principal points and foci, and the centres (nodal points) of Thomas Young form the basis of geometrical optics. By using several of the fundamental properties of double-ratios, properties that can be easily understood and applied by young students of average mathematical ability, the Gauss and Young points can be easily obtained by simple Geometry, and for direct and oblique incidence. The thick lens would be treated first. The reduction to thin lenses requires a little infinitesimal geometry, like the orthodox analytical method. Prisms, usually neglected, can also be dealt with. The theory should be accompanied by a well-systematised laboratory course to test the results experimentally, and a well-organised exercise class, that would include especially drawing with instruments on an adequate scale the more important loci of object and image. This can be easily and accurately done and proves of great interest as a survey and exploration of the optical field. Such drawings to

a pupil with a "geometric eye" have as much force and flexibility when effectively practised as models and experiments, indeed they are experiments in descriptive geometry. Optical instruments, spherical and chromatic aberrations, and the composition of individual sets of cardinal points into a resultant set for the whole system may also be displayed. The translation of the geometrical results into algebraic formulae either of the Newtonian or reciprocal type may be done at the end of the exercise. Thomas Young, in one of his writings—I regret that at the moment I cannot give the reference—stated very emphatically that for real enlightenment purely geometrical procedures were superior to analytical. This may have been associated with his paper on a centre for oblique refraction at a spherical surface, a centre that was adopted later by Cornu and Weierstrass. May I refer to papers by me in the Society's *Proceedings**, in which some of these matters are crudely discussed.

Reply by Mr T. Smith: It is obviously unnecessary for me to attempt a general summing up of the evidence that has been brought forward in the course of this discussion, but I would like to refer to two or three points of importance that have arisen. We all know that no teacher can afford to disregard the examination syllabus and neglect the kind of question that is set, but in spite of the temporary difficulty thus caused I believe that many teachers will sympathise strongly with Mr Vernon's emphasis on the cultural value of the instruction to be given. In this respect the geometrical treatment, in my judgment, inevitably fails, though, as indicated by Mr Vernon and Mr Emsley, it can be employed effectively in revising work originally treated on the wave basis. Mr Vernon's difficulty about sign conventions can be overcome by interchanging the object and image spaces. Mr Tarrant's objection about waves or quanta involves no difficulty, for we are always dealing with statistical resultants, and the simple wave theory is sufficient.

It would be easy to read into Mr Vinycomb's interesting figures more than they are entitled to convey. My experience in examining scholarship papers taught me that, with the exception of a rare student of outstanding ability, the quality of the teaching, rather than the knowledge and skill of the pupil, is being scrutinised. I would add, while mentioning examinations, that questions in terms of wave concepts, involving numerical answers that can be marked rapidly, may be set as readily as those now common: indeed most questions of practical interest can be dealt with more concisely by the wave treatment.

It has been suggested that the geometrical method has a decided advantage over the wave treatment in exhibiting the degree of approximation reached. The exact contrary is the fact. Moreover, it should be remembered that the physically significant quantity is the difference of path: other geometrical magnitudes may prove most misleading. As an extreme example an infinitely great longitudinal focal displacement (almost the only quantity considered in geometrical discussions) may correspond to a vanishingly small, and therefore physically imperceptible, difference of path.

* *Proc. Phys. Soc.* **25**, 160 (1913); **26**, 212 (1914).

Two points emerge unmistakably from this discussion. First, difficulties are experienced in the traditional presentation of the theory, and many teachers have given much thought to the way in which they can best be met. Secondly, judging from a scrutiny of the contributed papers, there is a decided majority in favour of adopting one particular course on some of the more debatable questions, such as the sign conventions. Since the existing divergencies cause considerable trouble to students, an attempt might well be made to see whether general agreement on these points can be reached, and I should like this meeting to express its views on the suggestion contained in Mr Williams' paper that the Council of the Physical Society should approach the Science Masters' Association and other interested bodies with a view to the appointment of a joint committee to make recommendations on these matters.

GENERAL SURVEY OF THE EFFECTS OF PRESSURE ON THE PROPERTIES OF MATTER

By Prof. P. W. BRIDGMAN, Hollis Professor of Mathematics and Natural Philosophy, Harvard University

The Fourteenth Guthrie Lecture, delivered April 26, 1929

Prof. Richardson has intimated that an acceptable subject for this lecture would be a general survey of my experiments on the properties of matter under high pressure, and I have accordingly made this choice. I feel, however, that this choice demands some apology on two counts: first, because little of what I have to say is new; and second, because the subject, concerned as it is with the properties of matter in bulk, is not to-day one of the most lively interest. There are, nevertheless, certain points of view from which such a topic may be regarded, on the contrary, as especially timely. If the Bohr-Heisenberg principle, with the radical change in our physical point of view which it involves, turns out to be correct, it will not be possible to make indefinite further progress in the direction of the analysis of the very small, and physics must soon return to the task of explaining the properties of matter in bulk, a task which has been temporarily laid aside. Furthermore, the Bohr-Heisenberg principle suggests that matter in bulk may have properties not deducible from the measurable properties of its smallest parts; if this is the case, it becomes doubly important to investigate the large scale properties of matter under those especially simple changes of condition produced by hydrostatic pressure. But in order that hydrostatic pressure may produce significant changes, it is necessary that it be of the same order of intensity as the atomic or molecular forces themselves, and these, for ordinary solids or liquids, are of the order of tens of thousands of kilograms per square centimetre. The point of the experiments upon which I am to report is that the pressures are of this order of magnitude.

This is not the place to go into questions of technique. It will be enough to say that by the use of a packing of special design, which automatically becomes tighter at higher pressures, the problem of leak disappears, so that the only limit to the pressure attainable is set by the cohesive strength of the walls of the containing vessels.* By the use of alloy steels, and of vessels of small size, so that the beneficial effects of heat treatment may be extended throughout the entire wall, pressures up to more than 20,000 kg./cm.^{2} can be handled and measured with an accuracy of 0·1 per cent. Most of my experiments, however, are made to only

* The way in which the automatic tightening of the packing is produced in the case of the plunger by which pressure is generated is shown in Fig. 1. This represents a section of a cylinder

12,000 kg., in the interests of economical life of the apparatus. There was, of course, a great deal of preliminary development work to be done in finding the best designs of the containing vessels, in developing methods of measuring pressure, and in devising means of measuring various properties of small amounts of substances enclosed in heavy steel vessels, but I leave this aspect of the subject to discuss the effects of pressure on various properties of matter.

Perhaps the most significant and certainly the simplest of the effects of pressure is diminution of volume. The change of volume of any truly homogeneous substance, free from internal strains, is entirely reversible with pressure; no permanent change of volume has ever been observed in such materials up to pressures at least as high as 25,000 kg. Contrary statements sometimes found in the literature are to be explained either by flaws in the material or else by failure of the pressure to be truly hydrostatic, as when pressure is transmitted by an oil which freezes under pressure, thus producing permanent changes of figure in the material.

in which pressure is generated in a liquid, *L*, by the advance of a piston of hardened steel, *P*, driven by a hydraulic press. The mushroom-shaped plug, *A*, carries a collar, *C*, of soft rubber packing. The plug with the packing is pushed forward by the piston *P* acting on the intermediary ring of hardened steel *D*. The space, *E*, at the end of the stem of the mushroom is empty, so that the

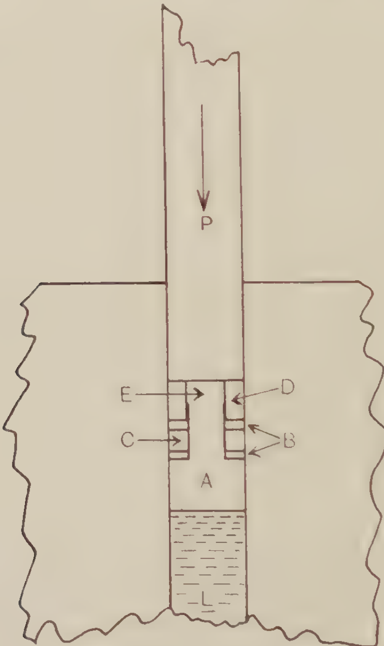


Fig. 1.

total pressure exerted by the ring of rubber must equal the total pressure exerted by the liquid on the head of *A*, and since the area of the ring is less than the area of the head, the pressure in lb./in.^{2} in the rubber is always greater than that in the liquid by the ratio of the areas, and the liquid can never leak. The packing is prevented from escaping by rings, *B*, of mild steel or copper. Further details of the technique are described in *Proc. Amer. Acad. Arts and Sci.*, **49**, 627–643 (1914).

The compressibility of fluids is, of course, in general much greater than that of solids. There is no essential distinction in compressibility between a substance ordinarily liquid and one of the so-called permanent gases, beyond the initial few thousand kilograms of pressure, which is far higher than the critical pressures between liquid and gas. It was shown by Amagat that air, for example, at a pressure of 3000 kg. is as dense as water.

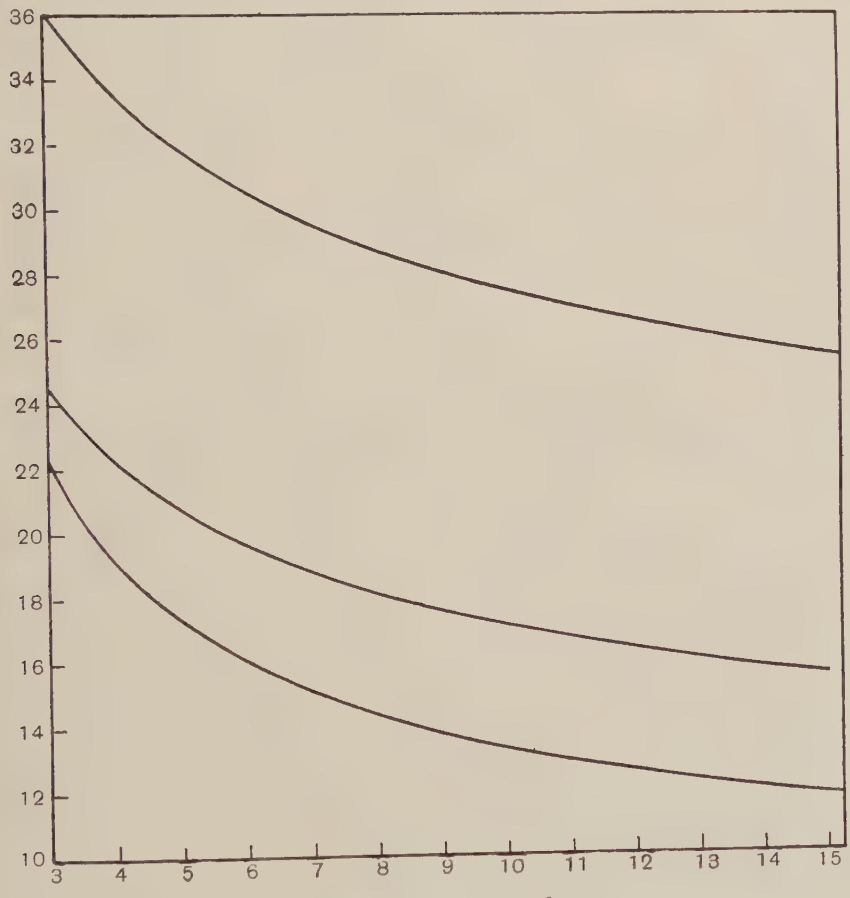


Fig. 2. The volume in c.c. per mol. as a function of pressure at 65° of nitrogen, hydrogen, and helium, reading from the top down.

The chief experimental difficulty in measuring the volume of gases at high pressures is the enormous initial volume, which makes it necessary to introduce the gas into the high pressure apparatus in small bombs in which it is subject to a high preliminary compression. In Fig. 2 is shown the volume in cm.^{3} per gm. molecule of N_2, H_2, and He between 3000 and 15,000 kg./cm.^{2}. The order of volumes is what is to be expected, monatomic He having a smaller volume at all pressures than biatomic H_2, which in turn has a smaller volume than N_2, also biatomic, but with

a much more complicated molecule. In Fig. 3 the product pv of these three gases is plotted after the manner of Amagat as a function of pressure. Departures of pv from unity are a rough measure of the departure of the behaviour of the substance

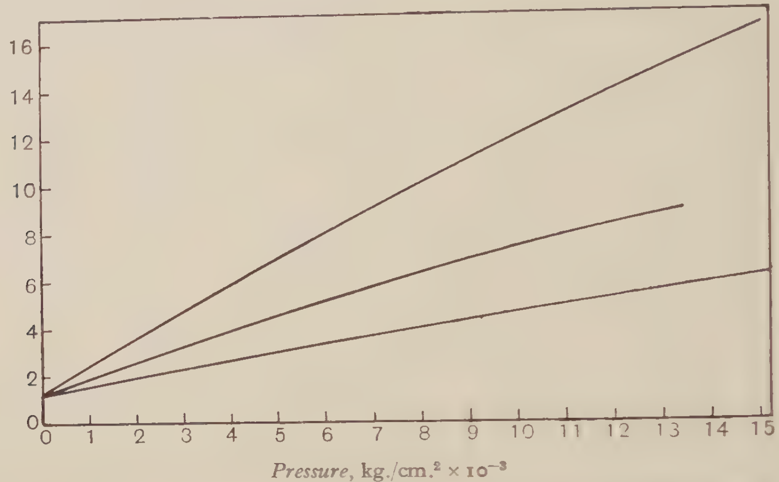


Fig. 3. The product pv as a function of pressure for nitrogen, hydrogen and helium (curves reading from the top down). p is the pressure in kg./cm.^{2} and v is the volume of that amount of gas which under a pressure of 1 kg./cm.^{2} occupies 1 c.c. at 0° C.

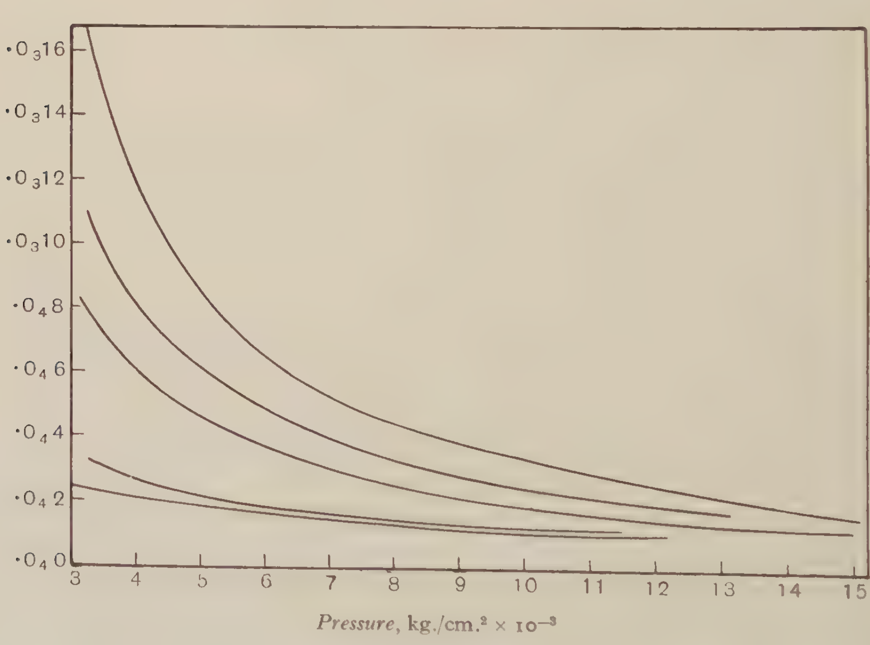


Fig. 4. The instantaneous compressibility, $\frac{1}{v}\left(\frac{\partial v}{\partial p}\right)_T$, as a function of pressure, of helium at 55°, hydrogen at 65°, nitrogen at 68°, CS_2 at 65°, and water at 65°, reading from the top down.

from that of a perfect gas. The departures found by Amagat up to 3000 kg. are seen to continue to become greater at very nearly a constant rate up to the highest pressure reached. Again, the behaviour is as is to be expected; He approaches most nearly to the perfect gas condition, although at 15,000 kg. its volume is six times greater than if it had remained a perfect gas. N_2 departs most widely from

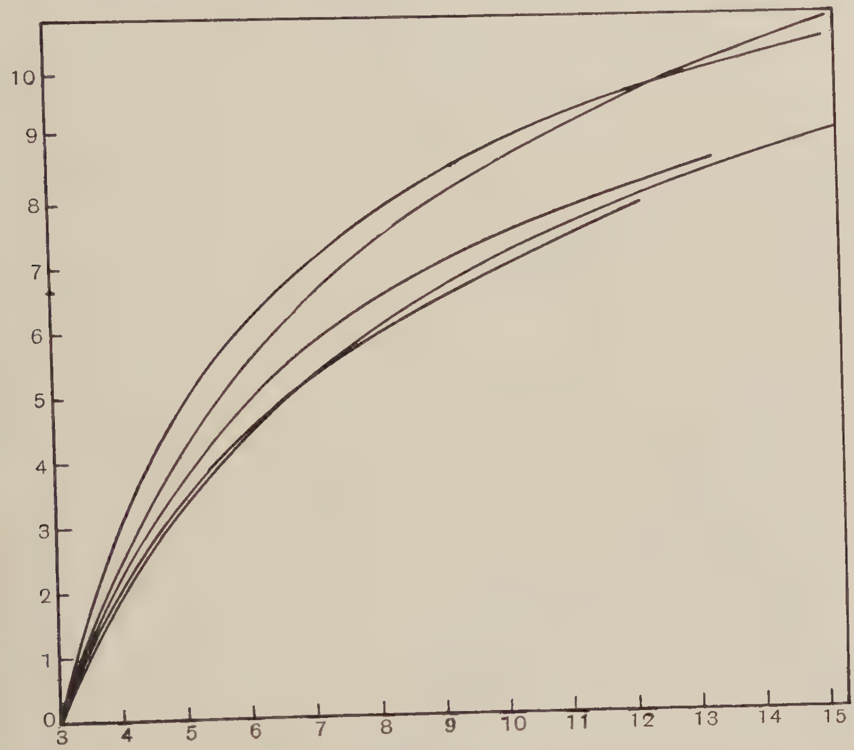


Fig. 5. The change of volume in cc. per mol as a function of pressure, reckoned from 3000 kg./cm.^{2} as the fiducial pressure. At 10,000 kg./cm.^{2} the order of the curves, reading from the top down, is: helium at 55°, nitrogen at 68°, hydrogen at 65°, argon at 55°, and ammonia at 30°.

the perfect gas, while H_2 is intermediate. In Fig. 4 the "instantaneous" compressibility, $\frac{1}{v}\left(\frac{\partial v}{\partial p}\right)_T$, of the three gases is shown and also, for purposes of comparison, the instantaneous compressibility of liquid CS_2 and water. At the lower pressures, in the neighbourhood of 3000 kg., the compressibility of the gases is several fold greater than that of the liquids, but at the highest pressure, 15,000 kg., the compressibility of all is nearly the same.

Thus far the behaviour of gases at high pressures has been a natural enough extension of the behaviour at low pressures. There are, however, effects at high pressure which would not naturally be inferred from the behaviour at low pressures. Fig. 5 shows one such effect, namely, the change of volume in cc. per mol, starting from the volume at 3000 kg. as the fiducial volume, for the five gases H_2, He, NH_3,

N_2 and A. The significant feature is the crossing of the curves for N_2 and He; it is evident, furthermore, that the curves for H_2 and A will also cross at a pressure slightly higher than that shown in the figure. The qualitative significance of this is as follows. At comparatively low pressures, the decrease of volume of a gas has its origin in a decrease of the empty space between the atoms or molecules, but as the molecules are pushed into closer contact, this effect becomes exhausted, and at high pressures this contribution to compressibility disappears. But there is another factor in the compression, namely, the actual loss of volume of the molecules themselves, and this evidently may persist at pressures where the initial effect no longer exists. This dual mechanism is doubtless the explanation of the striking difference between He and N_2. The atom of He is much smaller than the molecule of N_2, so that at low pressures the decrease of volume of He is greater than that of N_2. But at high pressures, where the important factor is the loss of volume of the atoms or molecules themselves, the decrease of volume of N_2 becomes greater than that of He, because the structure of the N_2 molecule is so much more complicated than that of the He atom that it has in it the potentiality of much greater loss of volume. The same sort of considerations also explain the relative behaviour of H_2 and A. The molecule of H_2 is normally smaller than the atom of A, so that at low pressures the compressibility of H_2 is greater than that of A. But at high pressures the rôles are reversed, because the electronic structure of atomic A is so much more complicated than the electronic structure of molecular H_2 that the atom of A is capable of much greater loss of volume than the molecule of H_2.

This behaviour, so definitely shown by these gases, is typical of the behaviour of all substances at high pressures. Beyond the first few thousand kilograms, the major part of the loss of volume is provided by the atoms or molecules themselves, and those substances with the most complicated or the most loosely constructed molecules have the greatest compressibility at high pressures, although at low pressures the behaviour may be the reverse. All this does not mean that a molecule is not properly to be regarded as a field of force rather than as a little nugget of matter in the old-fashioned sense, but it does mean that when the molecules are pushed closer together, there are qualitative changes in the interaction of their force fields similar to those which the older picture suggested.

Passing next to the volume behaviour of liquids, the most immediately striking fact is that the volume changes of many liquids tend to approach much more nearly to equality at high pressures than at low. Measurements on fourteen common liquids, including the first five alcohols, ether, CS_2, C_2H_5Cl and water, show an extreme variation in the loss of volume under 12,000 kg. from 21 per cent. for water, the least compressible, to 33 per cent. for ether, the most compressible, which thus under 12,000 kg. loses only 50 per cent. more volume than water, although its initial compressibility is four or five times greater. The compressibility of liquids drops very rapidly with increasing pressure as is shown in Fig. 6. At 12,000 kg. the compressibility of common organic liquids varies from 1/14th to 1/20th part of the initial compressibility; half of the drop to the final value is accomplished in the first thousand kilograms, and at 6000 kg. the drop is 95 per cent. completed.

Qualitatively, there is thus much similarity between liquids and gases. As in the case of gases, the initial high compressibility of liquids arises from the taking out of slack between the molecules, whereas the part of the compressibility which

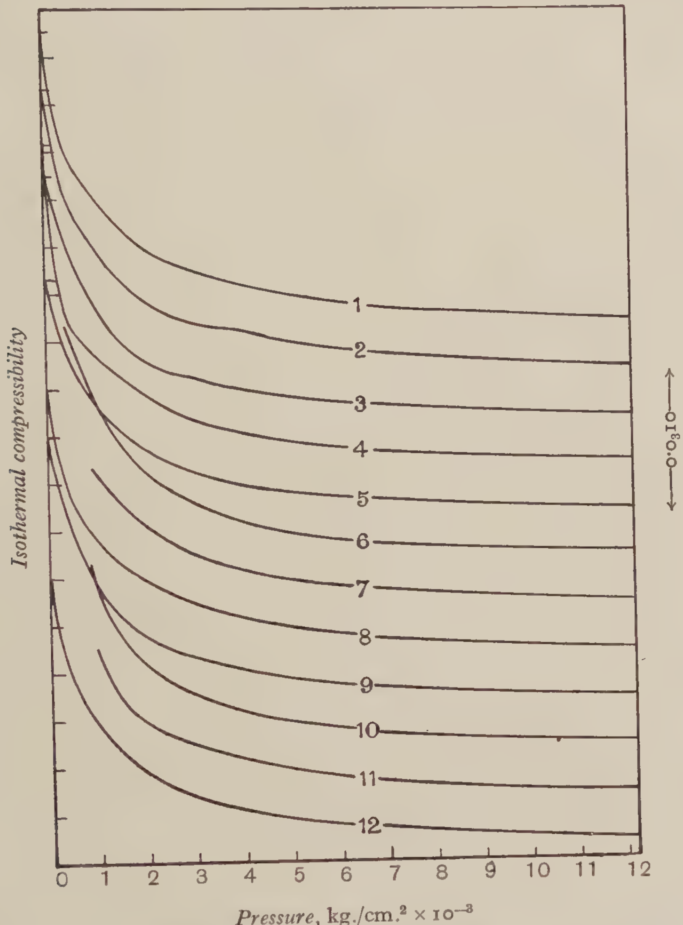


Fig. 6. The average compressibility between 20° and 80° of twelve liquids as a function of pressure. In order to prevent overlapping, the origin of each curve has been displaced one square with respect to the next. The scale of the curves is shown on the right hand side. The origin is so situated that the compressibility of each of the twelve liquids at 12,000 kg./cm.^{2} is between 0·0$_5$1 and 0. The numbers on the curves indicate the liquids as follows: 1, methyl alcohol; 2, ethyl alcohol; 3, propyl alcohol; 4, isobutyl alcohol; 5, amyl alcohol; 6, ether; 7, acetone; 8, CS_2; 9, PCl_3; 10, ethyl chloride; 11, ethyl bromide; and 12, ethyl iodide.

persists to high pressures probably arises from the decrease of volume of the molecules themselves. Liquid metals, as is to be expected, have a lower order of compressibility than organic liquids; it is also surprising that glycerine, which has a fairly complicated molecule, is only two-thirds as compressible as water at 12,000 kg.

By measuring the volume as a function of pressure at different temperatures it is possible to find how thermal expansion varies with pressure. The results for twelve different liquids are shown in Fig. 7. As was to be expected, thermal expansion drops with increasing pressure, but it is perhaps surprising that the decrease is much less than the decrease of compressibility; the thermal expansion under 12,000 kg. is on the average between one-fourth and one-fifth of its initial value, against four times as great a decrease in compressibility. Furthermore, the thermal

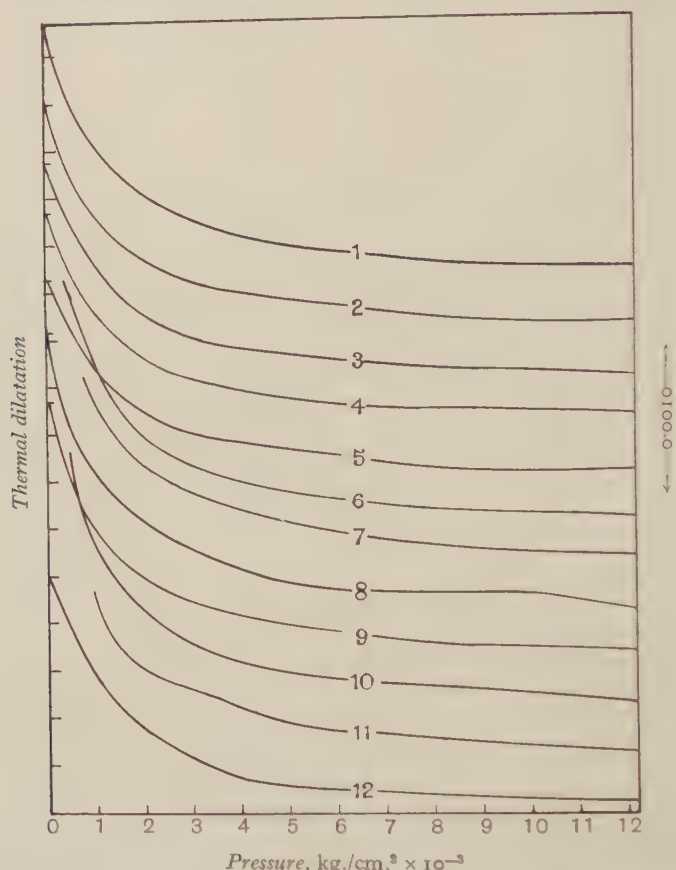


Fig. 7. The average thermal expansion between 20° and 80° of twelve liquids as a function of pressure. The general method of representation is the same as in Fig. 6, except that the origin for each curve is so situated that the expansion at 12,000 kg./cm.^{2} is between 0·0002 and 0·0003.

expansions of fourteen liquids, with the exception of water, approach much more closely to equality under 12,000 kg. than do the compressibilities, the extreme variation being by only 25 per cent., from 2·4 to 3·0 × 10^{-4}.

It is natural to expect parallelism between the effects of high pressure and low temperature, since both have a tendency to constrain freedom of internal motion, but the parallelism is far from complete, because as temperature approaches 0° abs., the compressibility drops only slightly, while the thermal expansion drops

to zero, but at high pressures the compressibility drops much more than the thermal expansion.

At atmospheric pressure the thermal expansion of all liquids increases with increasing temperature, but in the neighbourhood of 3000 kg. there is a reversal, and at higher pressures thermal expansion is greater at lower temperatures than at higher. The explanation is probably connected with the known fact that if the molecular restoring forces in a solid are linear functions of the molecular displacement, thermal expansion vanishes, so that a high thermal expansion means high departure from linearity. The smaller the volume of a substance the greater in general the departure of the forces from linearity, since the repulsive forces which predominate at small volumes vary inversely as some high power of the distance. At constant pressure the volume is less at low temperature, so that the departure from linearity would be expected to be greater, thus accounting for the greater thermal expansion. At low pressures there is a reversal of this behaviour because there is another mechanism active, the liquid tending to approach more nearly to a gas, with its high thermal expansion, the higher the temperature.

Superposed on the broad features just described, common to the behaviour of many liquids, there are small-scale specific differences of an almost indescribable complexity. An inspection of the experimental data would make it evident that an enormous number of parameters would be necessary to describe in full detail the behaviour of even a single liquid, and the small-scale complexities of different liquids are without discoverable relation to each other. This means that no such thing as a general equation of state for liquids can exist, and that the most that can be expected is to find an equation which shall reproduce the broad common features of behaviour described above. Very few attempts have been made in this direction, and as far as I know, none have been successful.

The volume behaviour of solids is qualitatively different from that of liquids or gases. In the first place, the compressibility is, as a general rule, less, as is to be expected. As far as I know, there is no exception to the rule that the compressibility of a substance in the solid phase is less than that of its own melt; this is true for normal substances which contract on freezing, and also true for water and bismuth which expand on freezing, and for which the contrary might be expected. It is thus evident that the lattice structure of itself imparts a certain stiffness foreign to the liquid. In the second place, the initial domain of high compressibility followed by very much lower compressibility, which is characteristic of liquids, is much less prominent in solids, the compressibility of which drops off comparatively little at high pressures. It is therefore probable that by far the larger part of the compression of a solid has its origin in the compression of its atoms. A striking example of this is caesium, the atom of which is highly complicated, and the compressibility of which is greater than that of any other metallic element. In Fig. 8 the volume of Cs is plotted against pressure and for comparison the volume of ether, the most compressible organic liquid. The decrease of volume of ether under the first thousand kilograms is nearly twice that of Cs, but the compressibility of Cs persists at high pressures, so that at 12,000 kg. the total volume decrement of

Cs has become very nearly as large as that of ether, and its actual compressibility is materially larger than that of ether.

As a general rule the compressibility of solids decreases at high pressures by amounts which are greater the greater the compressibility, as might be expected, but there are exceptions. The compressibility of pure quartz glass, SiO_2, and a number of compound glasses in which the content of SiO_2 is high, increases with increasing pressure by an amount far beyond experimental error. It has sometimes been thought that molecular stability demands that the compressibility becomes less as the volume becomes less, but this conclusion evidently rests on an incomplete analysis. I have recently found the same anomaly, that is, compressibility increasing with increasing pressure, in the element cerium. The behaviour of potas-

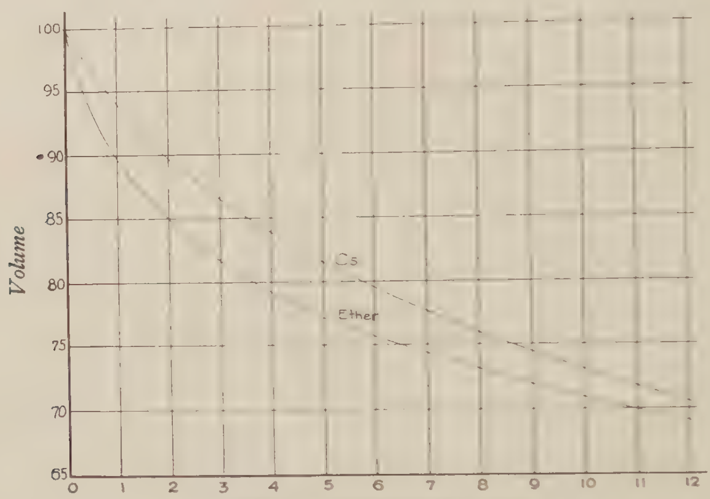


Fig. 8. The volume of ether and CS_2 as a function of pressure.

sium is also highly significant in this respect when compared with that of the other alkali metals. Fig. 9 shows that the compressibility of potassium drops off with increasing pressure much less than that of Rb or Cs, so that although initially the compressibility of Cs, for example, is nearly twice as great as that of K, at 12,000 kg. it has become considerably less. The persistence of compressibility of K is to be connected with the abnormally loose structure of the atom of K, which is shown by the fact that if the atomic volume of the alkali metals is plotted against atomic number, the volume of K is abnormally high. It is interesting, and probably of considerable significance, that at high pressures in the case of K there is a reversal in the direction of a simple function of the compressibility, which we do not need to explicitly define for our immediate purpose, of such a character that this function starts to approach the value which it would have if K were a perfect gas produced by the disintegration of its atoms into their component electrons and protons.

Of particular significance is the effect of pressure on single crystals. Only in cubic crystals is the change of volume under hydrostatic pressure the same in every

direction, so that to determine completely the effect of pressure on the volume of single crystals the linear compressibility must be measured in several directions. A large number of crystals have been studied from this point of view, and also the variation under pressure of electrical properties in different directions has been measured. The results are too numerous and complex to attempt to summarise here, except to state that the linear compressibility may vary much more with direction than might be expected from the difference of atomic spacing in different directions. Thus the linear compressibility of Zn parallel to the hexagonal axis is seven times as great as at right angles, whereas the ratio of the atomic spacing in the two directions is only 7 per cent. different from that for spherical atoms in normal hexagonal piling, and in fact the spacing is compressed along the axis,

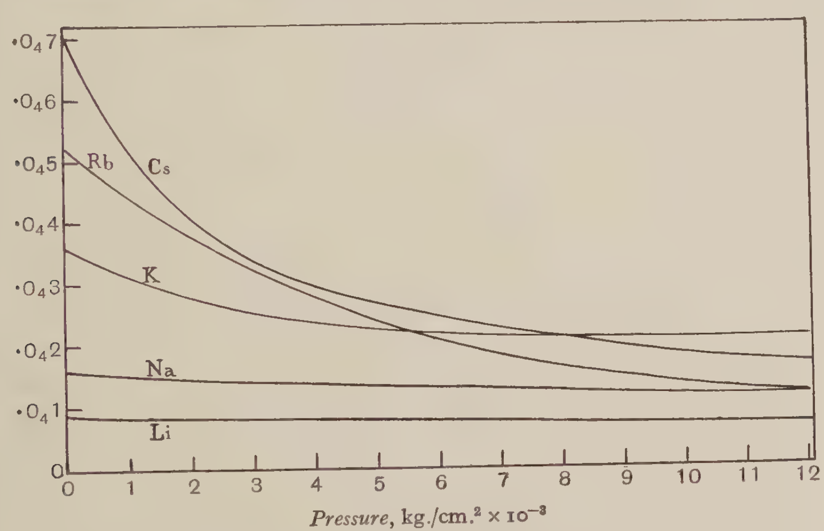


Fig. 9. The instantaneous compressibilities, $\frac{1}{v}\left(\frac{\partial v}{\partial p}\right)_T$, at 0° C. of the five alkali metals as a function of pressure.

while the compressibility is greater along the axis. The behaviour of tellurium is highly unusual, in that there is negative compressibility along the trigonal axis; that is, when exposed to hydrostatic pressure a crystal of Te expands in the direction of the axis. It is difficult to believe that effects of this kind can be explained in terms of atoms with spherical symmetry.

Perhaps the simplest of all crystals are the alkali halides. The compressibility of most of these has been measured by Slater up to pressures so high that good values could be found for the change of compressibility with pressure. This effect is at present beyond the reach of theory; although several theories have been proposed which give fairly good values for the initial compressibility, they all give changes of compressibility with pressure which are wide of the mark.

Next in simplicity after changes of volume come perhaps the changes of state produced by high pressure. We have to consider only the phenomena of change

from the liquid to the solid or from one solid to another, for the critical pressures between liquid and gas are of the order of a few hundred kilograms, and therefore

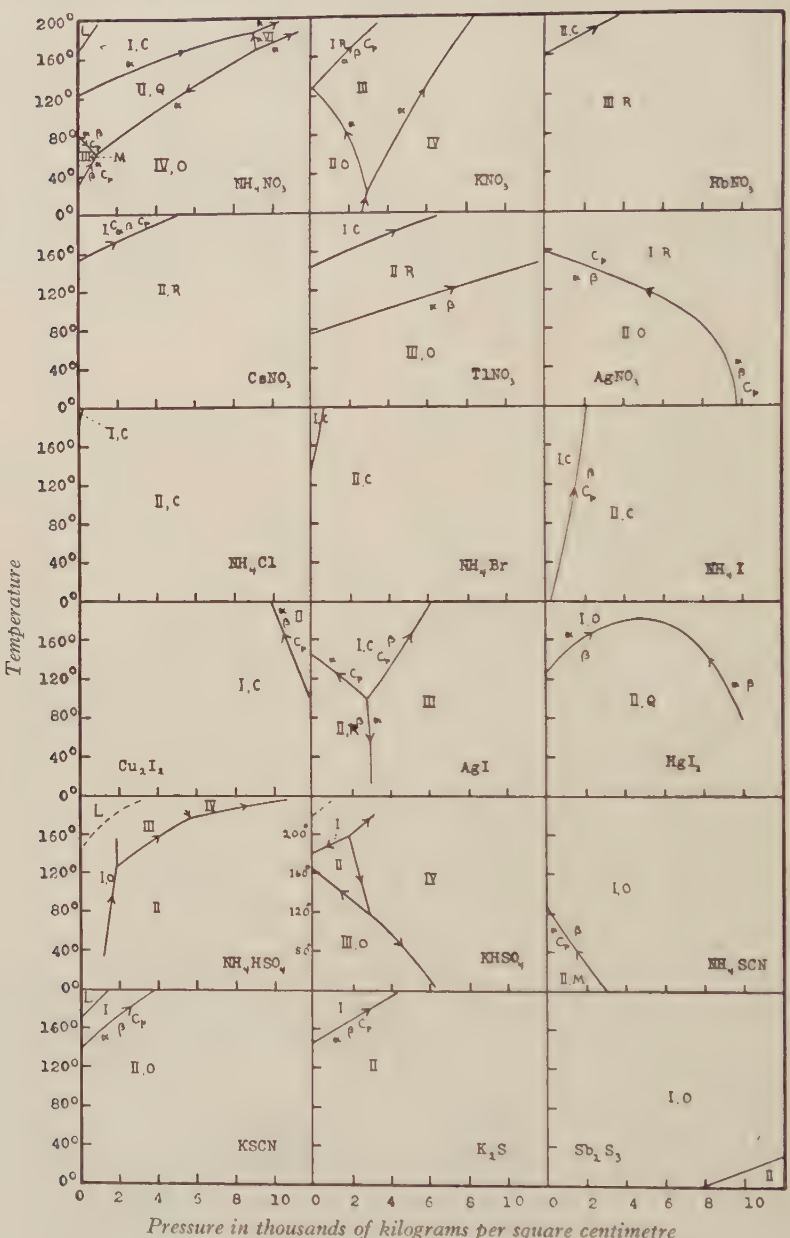


Fig. 10. The phase diagrams of a number of substances, showing the polymorphic transitions under pressure.

much lower than the pressures of interest here. There have been various ideas as to the ultimate relation between liquid and solid at high pressures. The most

natural of these, based on pure analogy with the critical phenomena between liquid and gas, was that there is a similar critical point between liquid and solid. Another idea prominently advocated by Tammann was that there is a maximum melting temperature, above which a substance is capable of existence only in the liquid phase, no matter how high the pressure. Experiment, however, seems favourable to neither of these points of view, for measurements on nearly forty substances up to 12,000 kg. and on water up to 21,000 kg. show that if either of these two possibilities ever occur, there must be a complete reversal of the universal trends in the experimental range of pressure. All melting curves, whether rising as is normal, or falling like those of water and bismuth, are concave toward the pressure axis. Furthermore, the difference of volume between liquid and solid plotted against melting pressure gives a curve convex toward the pressure axis, and the slope, dt/dp, of the melting curve plotted against temperature, gives a curve convex toward the temperature axis. A little consideration will show that these universal features demand that there can be neither critical point nor maximum. So far as I can see, there is no reason to think that the melting temperature may not be raised indefinitely by the application of sufficiently great pressure, but the rate at which the temperature is raised by a definite increment of pressure becomes continually less as the pressure becomes higher.

The phenomena of polymorphic transition from one crystalline phase to another show, on the other hand, no such regularities. A number of examples of the phase diagrams of polymorphic substances are shown in Figs. 10 and 11. One of the most interesting of these is that of water which is capable of existence in at least five solid forms under the proper conditions of temperature and pressure. The modifications called II and III in the diagram were first found by Tammann. Of the five modifications, ordinary ice is the only one which is less dense than the liquid, and therefore the only one whose melting temperature is depressed by increasing pressure. Camphor, with six modifications, is the only substance hitherto investigated with more modifications than water. NH_4NO_3 has five modifications, like water, in the ordinary temperature range, but at low temperatures there is another modification which there is some reason to think is merely one of the high temperature modifications reappearing at a lower temperature. Inspection shows that there is not the slightest similarity between the diagrams of the different substances. In general the phase diagrams of substances which are closely related are much less similar than their chemical similarity might lead one to expect. The diagrams of CCl_4 and CBr_4 are a case in point.

The only generalisation that can be made with regard to all these polymorphic changes is that there seems never to be a critical point between two different crystal modifications. This means that continuous transition from one space lattice to another never occurs, as seems natural enough. Apart from this, the most varied sorts of behaviour are possible. There are a comparatively large number of transition curves of the ice type, that is, those in which the modification stable at the higher temperature has the smaller volume, so that the transition temperature decreases with rising pressure. The transition curves may be either concave or

convex toward the pressure axis. They may have vertical tangents with either maximum or minimum pressures, and one example is known (HgI_2) of a horizontal tangent with maximum temperature. The mere existence of curves with horizontal or vertical tangents is of significance. At a horizontal tangent, the transition occurs with no volume change but with a finite latent heat; on the high-pressure

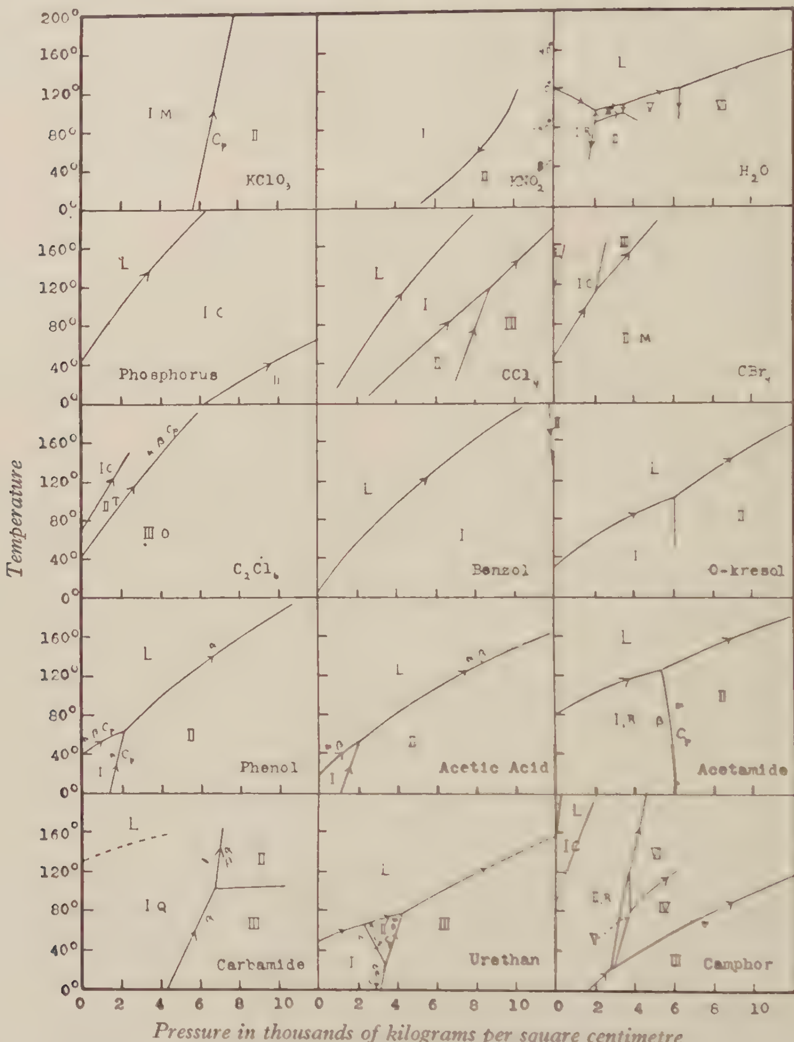


Fig. 11. More phase diagrams, similar to those of Fig. 10.

side of such a horizontal tangent the phase which has the larger volume has the smaller compressibility, a somewhat paradoxical effect. Similarly, at a vertical tangent the transition occurs with no thermal effect, but a finite change of volume; on the low-temperature side of such a vertical tangent the phase stable at the higher temperature has the lower specific heat, again paradoxical.

In addition to the information shown in the diagrams I have in a number of cases been able to measure the difference of compressibility between different

polymorphic forms of the same substance. The surprising result is that in the majority of cases that one of two phases which has the larger volume and which therefore is stable at the lower pressure is the less compressible. It is also true in general that the phase of larger volume has the higher crystalline symmetry. It seems as if the molecules or atoms have some internal structure analogous in some respects to a framework of high symmetry and rigidity, such that at low pressures, where the intensity of the external forces is low enough to permit that the natural arrangement be assumed, the molecules take an arrangement of high symmetry and low compressibility, with the projecting parts of the frameworks of different atoms in register with each other, but at high pressures the natural forces are overcome, the frameworks are pushed out of register, and the system collapses to an arrangement of smaller volume, of lower symmetry, and, because the frameworks are the most rigid part of the molecule, of greater compressibility. I doubt whether the complicated facts of polymorphism, or of the volume relations in liquid, can ever be explained in terms of simple central forces between atoms, although such an explanation is all that has hitherto been attempted, and in spite of the partial justification which such an explanation has lately received from the results of wave mechanics applied to simple systems.

It would appear, then, that a study of polymorphic behaviour gives a very delicate method of analysing the properties of matter in bulk, but at present the complications in the way of a theory are rather formidable. A systematic study of this field should start with the simplest systems. Among the series of substances which have been measured under pressure there are two comparatively simple. One of these is the series NH_4Cl, NH_4Br, NH_4I; these experience a polymorphic transition with very large volume change, and the lattice structure of the two modifications is known from X-ray studies at atmospheric pressure. The other of the series is RbCl, RbBr, RbI; the pressure of the transition is in the neighbourhood of 5000 kg., so that a direct X-ray determination of the lattice structure of the high pressure modification has not as yet been possible, but other considerations make it very probable that it is the same body centred cubic type as CsCl. No adequate explanation has yet been given of the occurrence of these two types of lattice in the alkali halide series, but enough has been done to show that the type of lattice which has the minimum free energy may be very sensitive to slight changes in the character of the interatomic forces.

Consider next the effect of pressure on electrical resistance. Out of 48 pure metals which have been measured, the resistance of 39 decreases under hydrostatic pressure by amounts varying with the character of the metal from 1 per cent. under 12,000 kg. for Co to 73 per cent. for K. The relation between change of resistance and pressure is not linear, but always the curve of resistance against pressure is convex toward the pressure axis, which means that the effect of a given increment of pressure becomes less at high pressures, as seems natural. In the case of K the initial rate of change is so great that resistance would entirely vanish at a pressure of 5600 kg. if it continued at the initial linear rate. Six or seven of the 48 metals, Li, Ca, Sr, Sb, Bi, the low-pressure modification of Ce, and perhaps Ti, increase in resistance under pressure, and in all these cases the curve of resistance against

pressure is convex toward the pressure axis, just as in the case of the metals with negative coefficient. But now the significance of convexity is that the effect of a given increment of pressure becomes greater at high pressures, which is not to be expected. It is particularly to be noticed that the curvature is such for these two types that the curve for one could be regarded as a prolongation of that of the other, so that it is natural to ask whether the resistance of all metals does not ultimately increase at sufficiently high pressures, the initial difference between different metals being ascribable to different internal pressures. Such an expectation is much strengthened by the discovery of three metals whose resistance does actually pass through a minimum with increasing pressure. The first of these is Cs, which is the most compressible metal, and therefore might be expected to show the effect most easily. The resistance of this at room temperature passes through a minimum of about 0·71 of the initial value between 4000 and 5000 kg. The next is Ba, which stands to the right of Cs in the periodic table, and which has a minimum at 0·97 of the initial resistance at about 8000 kg. Very recently I have found that Rb, which stands above Cs in the periodic table, also has a minimum at 18,000 kg. at 31 per cent. of its initial resistance. Measurements on K show no minimum up to 19,000 kg., but an easy extrapolation by first differences indicates pretty certainly a minimum in the neighbourhood of 24,000 kg. at 18 per cent. of its initial resistance. A similar, but much more uncertain extrapolation of the measurements on Na up to only 12,000 kg. is not unfavourable to the existence of a minimum, but indicates that it occurs at pressures probably considerably higher than 24,000 kg. However, on taking the last step in the series of the alkali metals, all regularity disappears, for the resistance of Li increases from the start with increase of pressure. In Fig. 12 the resistance under pressure of the alkali metals is shown. Furthermore, on passing across to the second column in the periodic table, the irregularities are more striking than the regularities. The resistance of Be decreases under pressure against the increase of Li, Mg decreases as does also Na, but Ca increases against the decrease of K, Sr increases against the decrease of Rb and its increase is greater than that of any other metal, while finally Ba has a minimum as does also Cs, but the whole scale of the pressure effects on Ba is very much smaller than on either Cs or Sr.

It is thus evident that in spite of the extreme simplicity of the change in the lattice structure produced by pressure, the effects of pressure on resistance are complicated, much more complicated than the effects of temperature, for example. I do not believe that any adequate explanation has yet been offered of the effects of pressure on resistance, even by the new wave mechanics. A consideration of the pressure effects makes it fairly certain, however, that the mechanism must be much more complicated than the pure temperature effects would lead us to expect. Particularly, the occurrence of a minimum resistance is suggestive; nearly always a minimum property involves the action of at least two mechanisms, which to a certain extent play against each other.

In addition to the effect of pressure on the electrical conductivity of pure metals, its effect on a number of binary alloys of metals has been measured. Examples are rather frequent here of a positive pressure coefficient of resistance.

One universal generalisation applies to all the measurements yet made on the pressure coefficient of alloys, both by other observers and myself, namely, that the initial effect of adding a small amount of foreign metal to a pure metal is to increase algebraically the pressure coefficient. That is, if the coefficient is initially negative, as it is in the majority of cases, adding a foreign metal makes the coefficient smaller numerically, while if the coefficient is initially positive, it becomes still more positive on addition of another metal. This is capable of a simple geometrical interpretation consistent with the wave mechanics picture of electrical resistance as arising from the scattering of electron waves by irregularities in the atomic structure. For, in addition to the irregularities in the lattice structure of a pure metal

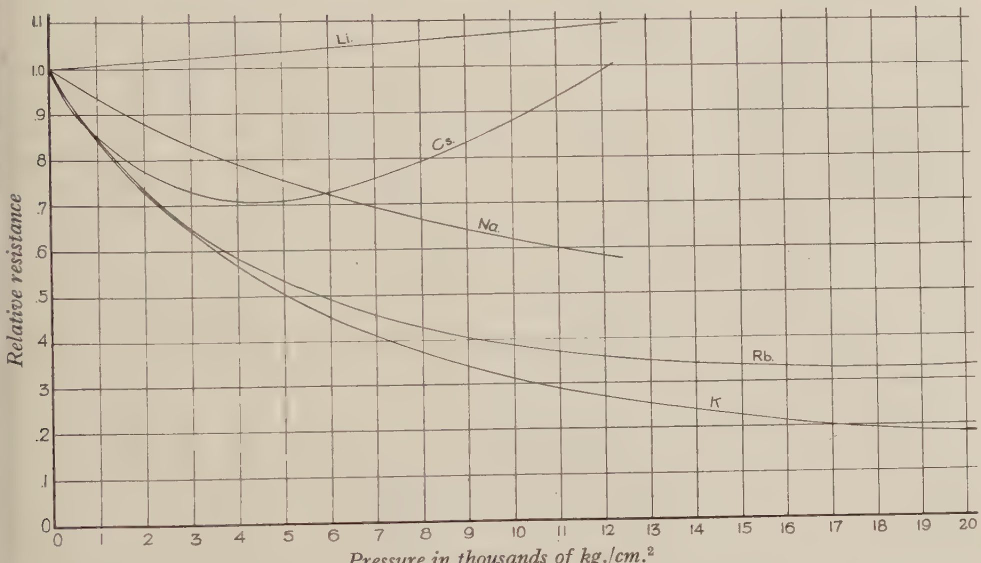


Fig. 12. The relative resistances under pressure of the five alkali metals.

due to temperature agitation, there must in general be irregularity of a purely geometrical character when a foreign metal is introduced, arising from the unequal sizes of the two sorts of atoms. The geometrical discrepancy between the two different sorts of atoms is accentuated as they are pushed into closer contact by pressure, so that the scattering of the electron waves arising from this effect increases, and the pressure coefficient becomes more strongly positive.

The thermal conductivity of metals is also recognised to be an electrical phenomenon, as proved by the universal value of the Wiedemann-Franz ratio. The effect of pressure up to 12,000 kg. has been measured on the thermal conductivity of 11 metals. The measurement is much more difficult to make than of resistance, and the results are not so accurate. Of the 11 metals, the thermal conductivity of 5 increases under pressure and that of 6 decreases. In only two cases does the Wiedemann-Franz ratio increase under pressure, and in the remaining 9 cases it decreases by amounts varying up to 15 per cent. for Ni at

12,000 kg. Since the Wiedemann-Franz ratio varies with pressure, it is obvious that the connection between electrical and thermal conductivity cannot be the universal and simple one of the original theory of Drude. A consideration of the contribution to the thermal conductivity by the atoms as distinguished from the electrons seems to offer a possibility of explaining at least part of the discrepancy, and would suggest that the contribution made by the atoms may be larger than often supposed.

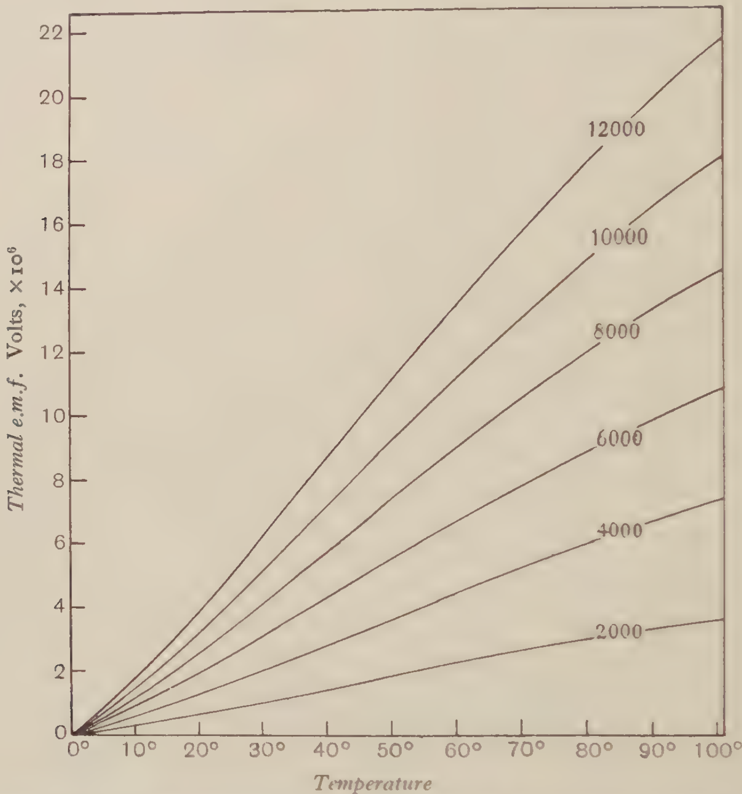


Fig. 13. Thermal e.m.f. of a couple composed of one branch of uncompressed pure platinum, with the other branch composed of the same metal compressed to the pressure in kg./cm.² indicated on the curves, the junctions being at 0° C. and the temperature plotted as abscissa.

Thermal e.m.f. is another electrical property that can be measured under pressure without too great difficulty. The effects are not small, for the thermal e.m.f. of a couple composed of a metal in the uncompressed state and the same metal compressed approaches the order of magnitude of the thermal e.m.f. of couples composed of ordinary dissimilar metals. For example, the thermal e.m.f. of a couple composed of Zn uncompressed and compressed to 12,000 kg., with its two junctions at 0° and 100°, is one-third as great as that of a couple of uncompressed Zn and Pb between the same temperature limits.

The effects of pressure on thermal e.m.f. may be exceedingly complicated and vary greatly from one metal to another, as Figs. 13 and 14 show. In general the

effect of pressure on Peltier and Thomson heats is positive, that is, the positive current absorbs heat in flowing from uncompressed to compressed metal, and the Thomson heat absorbed in passing from a low to a high temperature is greater in the compressed than in the uncompressed metal, but the irregularities are much greater than the irregularities in the resistance effects and occur in unexpected places, the irregularity of thermal e.m.f. being greatest for Sn, Fe, and Al, metals without special distinction in other respects under pressure. Just as in the case of resistance phenomena, these complicated effects indicate a complicated mechanism, certainly more complicated than contemplated in any theory yet proposed. There is no discoverable parallelism between the effects of pressure on resistance and on

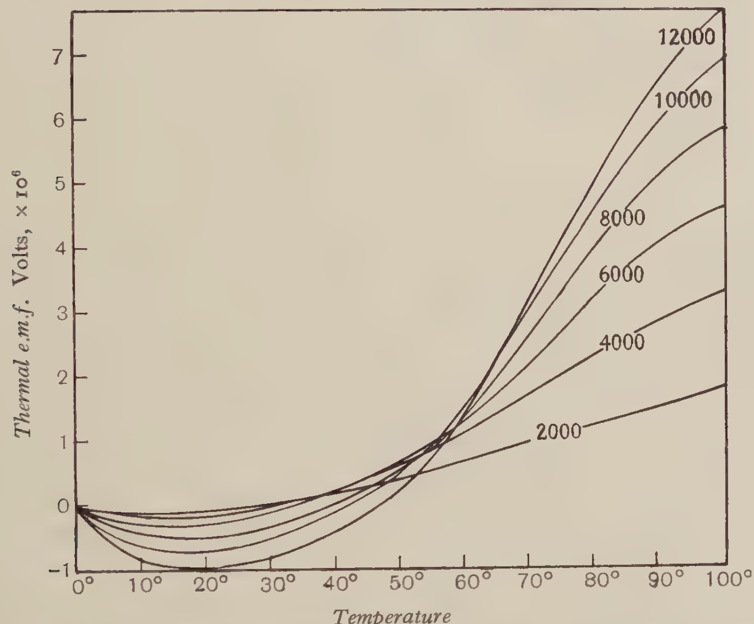


Fig. 14. Thermal e.m.f. of a couple composed of one branch of uncompressed pure iron, with the other branch composed of the same metal compressed to the pressure in kg./cm.² indicated on the curves, the junctions being at 0° C. and the temperature plotted as abscissae.

thermal e.m.f., from which one may perhaps draw the conclusion that essentially different aspects of the electron mechanism are responsible for thermoelectric phenomena, and the phenomena of resistance.

Passing now to another group of phenomena, the thermal conductivity of 15 liquids has been measured to 12,000 kg. The effect is to increase thermal conductivity by fairly large amounts, the factor of increase varying from 1·50 for water, the least compressible of the substances measured, to 2·74 for normal pentane, approximately the most compressible. It is very suggestive that this increase runs roughly parallel with the increase in the velocity of sound under pressure, as calculated from the change of density and the compressibility. I have found a very simple relation, suggested by this fact, for the thermal conductivity of liquids, namely, $k = 2\alpha v\delta^{-2}$. Here k is thermal conductivity in absolute units, α is the

molecular gas constant $2{\cdot}02 \times 10^{-16}$, v the velocity of sound, and δ the mean distance of separation of the molecules in the liquid, assuming them piled in simple cubic array. The formula represents with considerable success the thermal conductivity of normal liquids, of water, which is abnormal in so many respects, and also of non-crystalline solids, such as glass or hard rubber.

In gases the phenomena of viscosity are closely connected with those of thermal conductivity, but in liquids the difference in behaviour of the temperature coefficients of thermal conductivity and viscosity makes it highly probable that the mechanisms of the two effects are different. This view receives strong support from the pressure effect on viscosity, which has been measured for 43 liquids. In all cases, except the abnormal one of water and then only over a limited range of pressure and temperature, viscosity increases under pressure, and by amounts varying enormously from substance to substance. The factor of variation under 12,000 kg. is $1{\cdot}33$ for mercury, and 10^7 for eugenol ($C_3H_5.C_6H_3.OH.OCH_3$). In fact, the pressure effect on viscosity and its variation from substance to substance is much greater than any other known pressure effect. Viscosity increases geometrically with pressure, that is, the logarithm of viscosity is approximately a linear function of pressure. There is a very close correlation between the magnitude of the pressure effect and the complexity of the molecular structure, the pressure coefficient being least for monatomic Hg and greatest for complicated organic substances like eugenol. It is evident that no purely kinetic mechanism of viscosity, such as we have in gases, is competent to explain such enormously large effects. There would seem to be little question that a large part of the viscosity of a liquid is purely mechanical in origin, arising from the jambing together or interlocking of the molecules. Such a mechanism is consistent with the very large pressure effects, and the enormous variations of this effect with molecular complexity.

Finally, I may mention an investigation not yet finished on the effect of pressure on the shearing modulus or rigidity of solids. Hitherto the only elastic constant which has been measured under pressure is compressibility, but there would be considerable interest in determining how pressure affects all the constants. Unfortunately this is a matter of great experimental difficulty, and up to the present I have been able to measure the effect of pressure only on the shearing modulus of several isotropic substances. The rigidity of Fe increases under pressure, as would be expected, the increase being about $2{\cdot}5$ per cent. for 12,000 kg. The rigidity of glass, on the other hand, decreases by amounts varying with the composition from $0{\cdot}7$ per cent. to 11 per cent. for 12,000 kg. The effect is largest for those glasses which have the abnormal increase of compressibility with pressure, as might be expected. By combining this change in the shearing modulus with the known changes in density, it may be calculated that the velocity of a wave of shear decreases under 12,000 kg. by amounts varying from 2 to 6 per cent., a fact of some possible geological interest. By combining the pressure coefficient of rigidity with the pressure coefficient of compressibility, the effect of pressure on Young's modulus may be found; this may increase or decrease under pressure depending on the composition of the glass, by amounts varying from $+\,2{\cdot}5$ per cent. to $-\,10$ per cent. under 12,000 kg.

THE SPECTRUM OF TREBLY-IONISED THALLIUM

By K. R. RAO, D.Sc., Imperial College of Science, London

Communicated by Prof. A. FOWLER, F.R.S., *February* 21, 1929.
Read and discussed May 10, 1929

ABSTRACT. Using the measurements made by Carroll of the vacuum spark spectrum of thallium, combinations of the triad of triplet and singlet F, D, P terms of the d^9p configuration of Tl IV with the deeper ^{3}D and ^{1}D of the d^9s state have been found. Justification of the scheme is sought by comparison with similarly constituted spectra.

§ 1. INTRODUCTION

While the arc spectrum of thallium has long been known to consist of a well-developed doublet system, investigations for series relationships among the enhanced lines in the spark have been few. The first attempt at the analysis of the spark lines was that of Carroll*, who made extensive measurements of the "hot spark" spectrum of the element and indicated the chief doublets of Tl III. In a previous paper† the present writer, working in collaboration with Prof. Narayan and Mr A. S. Rao, has shown that some of the more easily excitable lines of the thallium spark form a triplet system characterising the spectrum of Tl II. The resonance and second ionisation potentials were found to be 6·47 and 20·3 volts respectively.

The measurements of Carroll contain strong unclassified lines in the extreme ultra-violet and it is believed that many of these may be due to the higher stages of ionisation of the element, as the method of excitation used by Carroll was particularly favourable for their production. The present paper shows that some of these lines can be grouped into a triplet system involving the deeper levels of trebly-ionised thallium. Evidence of the correctness of the identification is afforded by comparison with related spectra.

§ 2. THEORETICAL CONSIDERATIONS

The trebly-ionised atom of thallium contains a group of ten electrons outside the inner completed shells. Recent work of Shenstone‡, Laporte and Lang§, McLennan‖ and others¶ on the spectra of Ni-, Pd- and Pt-like atoms shows that

* *Phil. Trans.* (*London*), A, **225**, 357 (1926).

† *Ind. Journ. Phys.* 2, 467 (1928). Later McLennan and others, *Phil. Trans. R.S.C.* Sect. 3, **22**, 241 (1928), and also Smith, *Proc. Nat. Acad. Sc.* **14**, 951 (1928), came to identical conclusions.

‡ (Cu II), *Phys. Rev.* **29**, 380 (1927); (Ag II), **31**, 317 (1928); also Mazumdar, *Ind. Journ. Phys.* (1928).

§ (Zn III), *Phys. Rev.* **30**, 378 (1927); (Ga IV and Ge V), *Phys. Rev.* **31**, 748 (1927).

‖ (Ag II, Cd III, Au II, Hg III), *Trans. R.S.C. Sect.* 3, **22**, 1, 45, 103 (1927) and **22**, 247 (1928).

¶ (Cd III, In IV), *Phys. Rev.* **31**, 776 (1927); (Sn V), *Proc. Nat. Acad. Sci.* **14**, 345 (1928).

in such a ten-electron system the important configurations producing the optical spectrum are those built upon the normal state of the preceding ion, which is a ^2D term arising from nine equivalent n_3 electrons. The terms from such a configuration are, according to Hund's theory, given in the following table:

Table 1: Scheme of terms

5_1	5_2	5_3	5_4	6_1	6_2	6_3	Term prefix	Terms
2	6	10					$5d$	^1S
2	6	9		1			$6s$	^3D ^1D
2	6	9			1		$6p$	^3F ^3D ^3P ^1F ^1D ^1P
2	6	9				1	$6d$	^3G ^3F ^3D ^3P ^3S ^1G ^1F ^1D ^1P ^1S

The dashing of the terms is found unnecessary as, according to the notation adopted by Prof. Fowler, the term prefix itself determines the combinations.

§ 3. RESULTS

The results of the analysis of Tl IV are presented in Table 2 where frequencies are calculated from Carroll's list of wave-lengths. Term values are arbitrary and are found by assuming $^3D_3$ to be zero. Only combinations between the triad of triplet and singlet F, D, P terms of the d^9p state with the deeper ^3D and ^1D of d^9s have been found in this investigation. The three possible combination lines of the $6p$ terms with the deepest $5d\,^1S$ are expected to lie far outside the region investigated by Carroll and could not be located.

Table 2: Multiplets of Tl IV

$6p$ \ $6s$	$^3D_3$ *3586* 0	$^3D_2$ *15028* 3586	$^3D_1$ *3049* 18614	$^1D_2$ 21663
$^3P_2$ = 72582 *19854*	72582 (7)	68996 (5)	[53968]	50901 (3) *d*?
P_1 = 92436 *16314*	—	88848 (4)	[73820]	70775 (5)
P_0 = 108750	—	—	90136 (4)	—
$^3F_4$ = 90942 *− 16161*	90942 (6)	—	—	—
F_3 = 74781 *20488*	74784 (4)	71195 (8)	—	53115 (5)
F_2 = 95269	[95270]	91684 (4)	76654 (3)	73606 (5)
$^3D_3$ = 97009 *− 5650*	97006 (3)	93416 (x)	—	75362 (0)
D_2 = 91359 *23722*	91353 (0)	87773 (5)	72747 (5)	69701 (6)
D_1 = 115081	—	[111497]	96469 (3)	93416 ($4 - x$)
$^1F_3$ = 109822	—	—	—	88159 (4)
$^1P_1$ = 110975	—	—	92363 (4)	89309 (2)
$^1D_2$ = 113168	—	—	94557 (2)	91500 (2)

[ ] Bracketed values are calculated; the lines are expected to be faint.

While there seems to be no doubt as to the identification of the $6p$ triplet terms, as succeeding pages will show, there may be some uncertainty with regard to the

singlets $6p\,^1F\,^1P\,^1D$. They have been located mainly by consideration of the estimated positions of the lines and the occurrence of the separation 3049, equal to $6s\,^3D_1-6s\,^1D_2$, and they lack further check as their combinations with $6s\,^3D_{32}$ fall far outside Carroll's list. The vacuum spark spectrum of thallium must be investigated below λ 900 before these levels can be established with certainty. The levels $^3P_0$ and $^3F_4$ had to be fixed only by a single transition in each case, but alternative choice* was found unsatisfactory. The line ν 93416 occurs twice in the table; the intensity of the line, however, supports the assignments. It is further significant to observe that the combination $4s\,^3D_2-4p\,^3D_3$ is either faint or absent in Ni I-like spectra.

An interesting point as regards the intensities of the lines deserves mention. While it may not be safe to conclude from visual estimates of intensities of such widely separated lines, there appears still to be in Tl IV a definite increase in intensities of the intercombination lines relative to those of triplet-triplet and singlet-singlet combinations. This feature was observed also by Mack, Laporte and Lang in Ga IV and Ge V.

§ 4. COMPARISON WITH RELATED SPECTRA

Apart from the recurrence of constant frequency-differences and the location of intercombination lines, a valuable check on the identification of the members is afforded by a comparison with spectra of related elements. Our knowledge of the spectra of Au II, Hg III and Ga IV, In IV permits of such a comparison.

Laporte and Lang have shown that the application of the relativity laws to iso-electronic spectra, successfully employed by Millikan and Bowen in the case of "normal" spectra, can be fairly utilised for the prediction and location of line groups which occur in spectra with a large number of electrons, and involve separations between sub-levels large compared with those between different terms.

The regular doublet sequence is shown in Table 3.

Table 3: Regular doublet sequence

(At. No.) Element	$6s\,^3D_3-6s\,^3D_1$ ($\Delta\nu$)	$\sqrt[4]{\frac{\Delta\nu}{0\cdot00776}}$	s	Δs
78 Pt I	10132	33·8	44·2	1·0
79 Au II	12728	35·8	43·2	0·8
80 Hg III	15556	37·6	42·4	0·7
81 Tl IV	18613	39·3	41·7	

The values of the screening constant (s) exhibit the familiar slow decrease with increasing atomic number.

The irregular doublet sequence is indicated in Table 4, where the mean of the frequencies of the lines in each multiplet is taken for purposes of comparison.

* ν 92618 (6) might be suggested as equally probable for $^3D_3-^3F_4$.

Table 4: Irregular doublet sequence

Element	$(6s\,^3D–6p\,^3F)$	Diff.	$(6s\,^3D–6p\,^3P)$	Diff.	$(6s\,^3D–6p\,^3D)$	Diff.
Au II	54004	14831	47494		57423	18464
Hg III	68835	14586	—		75887	17007
Tl IV	83421		74725		92894	

Table 5 gives the separations of the various terms in the iso-electronic spectra Au II, Hg III, Tl IV on the one hand and in the spectra of the elements Ga IV, In IV and Tl IV of the same chemical group on the other.

Table 5: Term intervals

Term	Au II	Hg III	Tl IV	In IV	Ga IV
$6s\,^3D_3$					
D_2	2602	3179	3586	2197	1455
D_1	10125	12377	15028	4912	2120
$6p\,^3P_2$					
P_1	10351	15057	19854	6658	3081
P_0	9210	—	16314	4395	1820
$^3F_4$					
F_3	− 7492	− 12366	− 16161	− 4452	− 1271
F_2	11656	15300	20488	8651	2358
$^3D_3$					
D_2	− 1613	− 3052	− 5650	− 3824	636
D_1	12529	17928	23722	9522	3080

A general variation of the intervals between the terms of these spectra is obvious. But, as is expected, the variation is more regular and progressive in the sequence of spectra Au II, Hg III, Tl IV, than in that of Ga IV, In IV, Tl IV.

The relative values of singlet and triplet terms of a given configuration present considerable interest. While in Ni- and Pd-like spectra, and also in Au II, $6p\,^3D_1$ is deeper than $6p\,^1D_2$, the reverse is the case in Hg III, and if the identification of $6p\,^1D_2$ in Tl IV proves correct, this feature recurs in this spectrum also. The partial inversion of $6p\,^3F$ and ^3D and the complete inversion of $6p\,^3P$ and $6s\,^3D$ are analogous to those of Au II and Hg III.

The ratio of the separations $\{(ms)\,^3D_2-{}^3D_3\}/\{^3D_1-{}^3D_3\}$ is shown in the case of these ten electron systems in Table 6, which is an extension of a table given by Mack, Laporte and Lang*. They have shown that the progressive decrease of this ratio from Ni I to Ge V and the marked decrease from Ag II to Sn V are evidences against the following correlation of limits by Hund:

$$d^9s\ ^3D_3\ ^1D_2 \rightarrow d^9\ ^2D_{5/2}$$
$$^3D_2\ ^3D_1 \rightarrow\ ^2D_{3/2}.$$

A similar marked decrease of this ratio occurs also in Au II to Tl IV. The variations of the levels, in these spectra, as the nuclear charge increases, indicates that the limits for the levels $^1D_2$ and $^3D_2$ should be interchanged.

* *loc. cit.* p. 763.

Table 6: Ratio of separations

Ni I 0·448	Cu II 0·444	Zn III 0·428	Ga IV 0·407	Ge V 0·384
Pd I 0·337	Ag II 0·345	Cd III 0·330	In IV 0·309	Sn V 0·287
Pt I 0·077	Au II 0·257	Hg III 0·204	Tl IV 0·193	Pb V —

Finally, the lines of thallium classified in this investigation are collected in Table 7 for the sake of convenience.

Table 7: List of the lines

λ (int.)	ν (vac.)	Classification
1964·6 (3)	50901	$6s\ {}^1D_2-6p\ {}^3P_2$
1882·7 (5)	53115	${}^1D_2-\ {}^3F_3$
—	[53968]	${}^3D_1-\ {}^3P_2$
1449·37 (5)	68996	${}^3D_2-\ {}^3P_2$
34·71 (6)	69701	${}^1D_2-\ {}^3D_2$
12·93 (5)	70775	${}^1D_2-\ {}^3P_1$
04·60 (8)	71195	${}^3D_2-\ {}^3F_3$
1377·75 (7)	72582	${}^3D_3-\ {}^3P_2$
74·62 (5)	72747	${}^3D_1-\ {}^3D_2$
58·58 (5)	73606	${}^1D_2-\ {}^3F_2$
—	[73820]	${}^3D_1-\ {}^3P_1$
37·19 (4)	74784	${}^3D_3-\ {}^3F_3$
26·93 (0)	75362	${}^1D_2-\ {}^3D_3$
04·55 (3)	76654	${}^3D_1-\ {}^3F_2$
1139·30 (5)	87773	${}^3D_2-\ {}^3D_2$
25·52 (4)	88848	${}^3D_2-\ {}^3P_1$
19·71 (2)	89309	${}^1D_2-\ {}^1P_1$
09·43 (4)	90136	${}^3D_1-\ {}^3P_0$
1099·60 (6)	90942	${}^3D_3-\ {}^3F_4$
94·65 (0)	91353	${}^3D_3-\ {}^3D_2$
92·90 (2)	91500	${}^1D_2-\ {}^1D_2$
90·70 (4)	91684	${}^3D_2-\ {}^3F_2$
70·48 (4)	93416	${}^3D_2-\ {}^3D_3$ also ${}^1D_2-{}^3D_1$
57·56 (2)	94557	${}^3D_1-\ {}^1D_2$
—	[95270]	${}^3D_3-\ {}^3F_2$
36·60 (3)	96469	${}^3D_1-\ {}^3D_1$
30·86 (3)	97006	${}^3D_3-\ {}^3D_3$
—	[111497]	${}^3D_2-\ {}^3D_1$

§ 5. ACKNOWLEDGMENT

I wish to express my grateful thanks to Prof. A. Fowler, F.R.S., for his stimulating interest in the work and helpful criticism.

DISCUSSION

Mr E. W. H. Selwyn: I think that the manner in which Mr Rao has sorted out wave numbers from a list of practically completely unclassified lines, until they represent a useful analysis, is much to be admired.

THE ELASTIC PROPERTIES OF THICK CYLINDRICAL SHELLS UNDER INTERNAL PRESSURE

By G. A. WEDGWOOD, M.Sc., East London College

Received February 26, 1929. *Read and discussed May* 10, 1929

ABSTRACT. An account is given of an experimental investigation of the usually accepted theory relating to the stresses in a thick cylindrical shell under pressure. The theory is examined in the light of the longitudinal and diametral extensions of a number of steel cylinders which were subjected internally to hydrostatic pressure, the cylinders being closed at the ends by covers secured to the shell itself. Formulae derived from the theory provide a ready means of arriving at the values of the elastic constants E, σ and K. The values of E and σ were found by pressure experiments and also by direct tension experiments in a lever-testing machine. The results show a discrepancy of about 3 per cent. between the values of the constants thus obtained. The longitudinal extensions in the pressure experiments should give the bulk modulus K, but the value thus obtained is about 7 per cent. less than the value obtained by calculation from the determinations of E and σ in direct tension experiments. The author endeavours to explain these discrepancies as being due to the non-isotropic nature of the material, and he shows how on this assumption the values of E and σ for a lateral direction may be estimated from his results.

The paper gives some information concerning the high-pressure apparatus which was employed, this being designed on the lines suggested by Prof. P. W. Bridgman; a comparison is also made between the working of the roller extensometer developed by Prof. E. H. Lamb and two other extensometers commonly used in engineering laboratories.

§ 1. SYMBOLS USED IN THE TEXT

p, pressure.
q, hoop stress.
f, longitudinal stress.
r, radius.
k, ratio of outer diameter to inner diameter of shell.
e_2, diametral strain.
e_3, longitudinal strain.
E, Young's modulus.
K, bulk modulus.
σ, Poisson's ratio.
d, mean of diameters of extensometer rollers in inches.
L, distance from scale to extensometer mirror facing it in cm.
l, distance between extensometer mirrors in cm.
x, extensometer scale reading in cm.

§ 2. INTRODUCTION

THE usually accepted theory* relating to the stresses in thick cylindrical shells under pressure leads to the equations

$$p = A + B/r^2, \quad q = A - B/r^2 \qquad \text{......(1)},$$

in which A and B are constants. The theory assumes a perfectly isotropic material, that a cross-section remains plane and generating lines straight after the deformation of the material has taken place, and that the alterations in the radial dimensions of the shell are inappreciable after the application of the pressure.

The values of the constants A and B are determined by the conditions at the inner and outer surfaces of the shell. If there is a hydrostatic pressure p_0 acting on the inner surface, whilst the outer surface is at zero pressure, we have

$$A = \frac{r_1^2}{r_2^2 - r_1^2} p_0, \quad B = -\frac{r_1^2 r_2^2}{r_2^2 - r_1^2} p_0 \qquad \text{......(2)},$$

in which r_1 and r_2 are the inner and outer radii of the shell respectively.

If the shell is closed at both ends, the internal pressure p_0 produces in the material a longitudinal stress f, and $f.\pi(r_2^2 - r_1^2) = p_0 \pi r_1^2$,

or $$f = \frac{r_1^2}{r_2^2 - r_1^2} p_0 \qquad \text{......(3)}.$$

But $$e_3 = f/E - (\sigma/E)(p + q) \qquad \text{......(4)},$$

and hence from (1), making use of (2) and (3), we have

$$\frac{E}{1 - 2\sigma} . e_3 = \frac{p_0}{k^2 - 1} \qquad \text{......(5)}.$$

Again, at the outer surface $$e_2 = q/E - \sigma f/E \qquad \text{......(6)}.$$

Putting in the values of q and f, we have

$$e_2 . E/(2 - \sigma) = p_0/(k^2 - 1) \qquad \text{......(7)}.$$

From (5) and (7), for the same value of p_0,

$$(2 - \sigma)/(1 - 2\sigma) = e_2/e_3 \qquad \text{......(8)}.$$

The formulae (5), (7), and (8), if the theory is correct, afford us a means of determining E and σ directly by extensional measurements.

The object of the present work is to examine the above theory by means of longitudinal and diametral extensional measurements of a thick cylindrical shell subjected internally to hydrostatic pressure, the cylinder being closed at both ends by covers secured to the shell itself. From these measurements the elastic constants E and σ are calculated according to the theory and are compared with the values determined by measurement of the longitudinal extensions and transverse contractions under tensile stress.

If work of this kind is to be carried out successfully, one must employ a material which from the standpoint of its elasticity is as nearly perfect as possible. It is also essential that the hollow cylinders be made with the greatest care with respect to their inner and outer surfaces being co-axial, and with respect to the generating lines of the surfaces being parallel.

* Love, *Mathematical Theory of Elasticity* (1928).

For the purpose of the measurement of the longitudinal and diametral strains one must employ extensometers which are designed to measure accurately very small extensions. The longitudinal strains, especially in the case of the thicker shells, are, as indicated by the above theory, very small, much smaller than the corresponding diametral strains. The longitudinal extensometer, while it must be a very sensitive instrument, must also be such that its precision in working is not affected by the disturbance which unavoidably occurs at its points of contact with the cylinder owing to the bulging of the shell.

Measurements by delicate instruments of this kind must of course be taken in a room free from sudden changes of temperature.

For the generation and accurate measurement of the hydrostatic pressure inside the shell special apparatus is necessary. The pressures are high, as great as 15 tons per square inch (2286 atmospheres) if steel cylinders are employed in the investigation. For the measurement of pressures of this magnitude apparatus capable of greater accuracy than a Bourdon pressure gauge is desirable.

§ 3. THE MATERIAL

The thick cylindrical shells were made of high quality steel from four bars supplied by Messrs Thos. Firth and Sons, and the cylinders in the course of making were normalised. The analysis of each bar is shown in Table 1.

Table 1: Analysis of metal bars

Designation of bar	C	Si	Mn	S	P	Ni	Cr
1	0·39	0·22	0·58	0·016	0·012	3·48	0·07
2	0·43	0·28	0·53	0·017	0·029	0·27	—
3	0·28	0·15	0·41	0·018	0·029	0·15	—
4	0·09	0·06	0·28	0·021	0·012	0·27	—

The ultimate tensile strength of these steels after normalising was between 22 and 50 tons per square inch.

Seventeen specimen cylindrical shells were made from the four bars. The shape of each specimen is shown in Fig. 1, and the dimensions, together with the value of $(k^2 - 1)$ for each, are given in Table 3.

In the making of a specimen every precaution was taken in order to ensure that the inner and outer surfaces were co-axial. The work of drilling the longer specimens was kindly undertaken by Messrs The Birmingham Small Arms Co.

§ 4. THE EXTENSOMETERS

The mirror extensometers, both longitudinal and lateral, developed by Prof. E. H. Lamb*, are capable of a very fine degree of accuracy, and would appear to be admirably suited to work of this nature. Certain slight modifications were made in the construction of these instruments, and in their method of use in the present investigation.

* *Engineering*, **119**, 207 (1925).

In order to measure the longitudinal extensions of the shorter cylinders under test, an instrument which could deal with the extensions of a 2-in. length of shell was made, the diameters of the rollers being about 0·0625 in. This instrument is seen in position on the thick cylinder under pressure in Fig. 1. It is here seen that between each of the two nuts and the clamp which secures the two parts of the

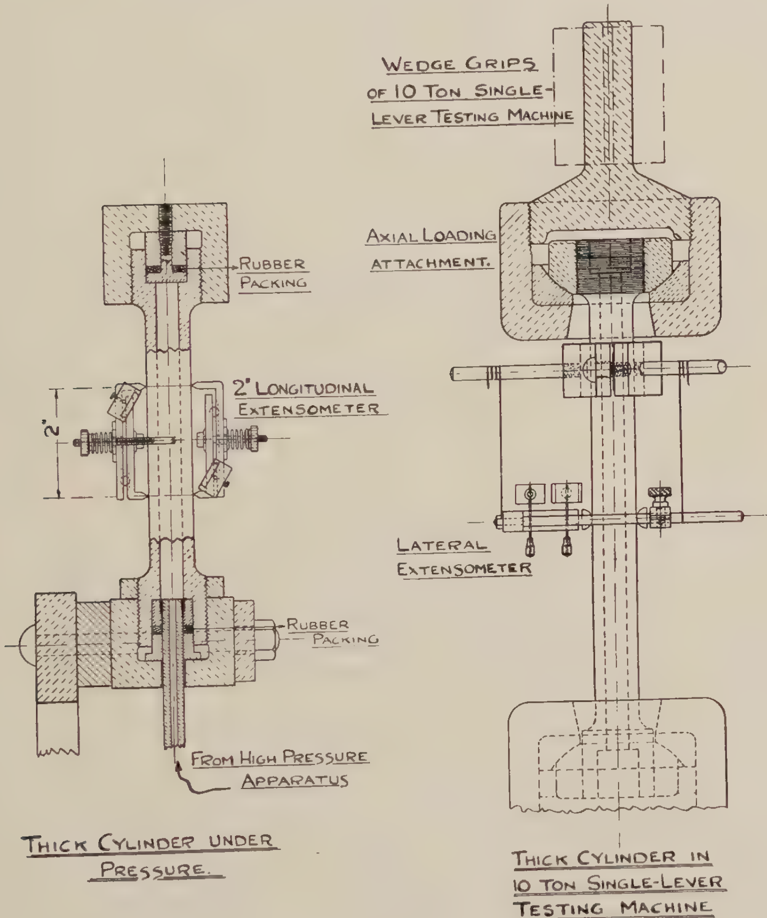


Fig. 1.

extensometer to the cylinder a light spring is inserted. This, by being flexible, takes up the bulge of the shell, and thus the tendency of the left-hand roller to slip rather than to roll between the pair of distance strips is minimised.

Each mirror-holder was balanced rotationally on its roller, and the knob at the other end of the roller was of about the same mass as the mirror-holder. Then, in order to be certain as to the exact distance, nominally 2 in., between the knife-edges, and also to be certain as to the grip of the instrument on the specimen, two fine grooves about 0·002 in. deep were scratched on the surface of the cylinder, and

particular care was taken when setting up the instrument to see that the four points of contact were in these grooves.

In the case of the lateral extensometer an attempt was made to obtain point contact between the instrument and the surface of the specimen by having the "contacts" of semi-cylindrical shape as shown in the right-hand view of the thick cylinder in the 10-ton testing machine (Fig. 1). The extensometer was also supported lightly by three pieces of thin cord. In the pressure experiments these were made fast to a plate attached to the top nut which closed the thick cylinder when under pressure.

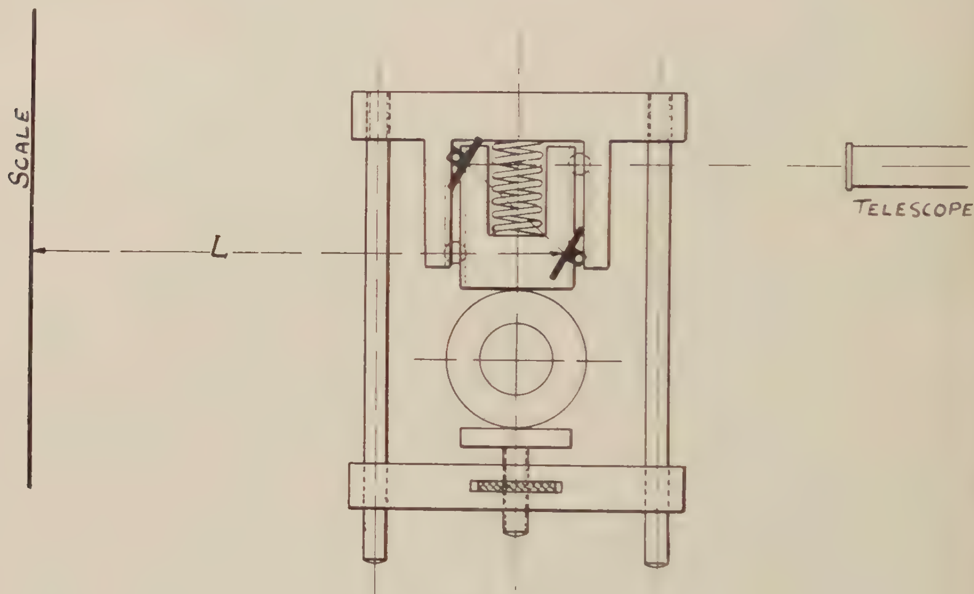


Fig. 2.

The geometry of the optical arrangement is the same for both instruments. Referring to Fig. 2, it is easily proved that the lateral extension of the specimen is equal to $\frac{d}{4L+2l}.x$, and this holds for the longitudinal extensions as measured by means of the 2-in. instrument shown on the left of Fig. 1. The value of d can be obtained accurately by means of a Newall measuring machine, the values of L and l by means of a tape measure, and the scale reading x, with the help of a good telescope provided with a proper spider-line, can be read accurately to the nearest 0·1 mm. at a distance of 460 cm.

Sudden temperature changes are always a serious difficulty in accurate extensometer work, and so is vibration. For these reasons the work here recorded was carried out in a cellar which was some distance away from machinery and was free from sunlight and draughts.

The effect of placing a screen between the observer and the extensometer was tested. It was found that the rate of rise in temperature, as recorded by a delicate

thermometer placed in the immediate vicinity of the specimen, with the screen in position was less than half what it was without the screen. But it was also found in the course of the investigations that the extensional measurements did not show any appreciable discrepancy between readings taken with and without the screen. The instrument is itself perhaps as good an indication of sudden temperature changes in the room as any sensitive thermometer.

§ 5. THE HIGH-PRESSURE APPARATUS

The high-pressure apparatus was designed and constructed on the lines suggested by Prof. P. W. Bridgman in his various contributions to the *Proceedings of the American Academy*, **44**, **47**, and **49**. It consists of a force-pump, a hydraulic press, and a cylinder in which the high pressures are generated and measured. Two sectional views of the hydraulic press are given in Fig. 3, and a sectional view of the high-pressure cylinder in Fig. 4, and two views of these parts assembled together are to be seen in Fig. 5. The force-pump was made by C. W. Cook, Ltd., and was designed to deliver liquid at pressures up to 5 tons per sq. in. The pump in communication with the $2\frac{1}{2}$-inches-diameter piston of the hydraulic press forces a

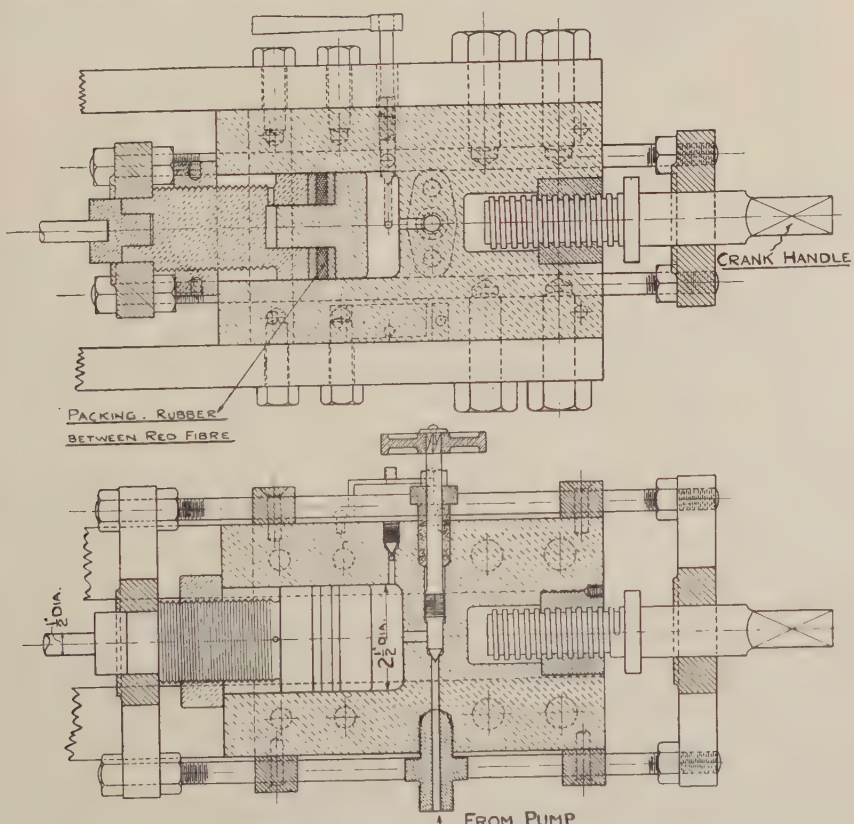


Fig. 3.

$\frac{1}{2}$-inch-diameter plunger into the high-pressure cylinder. Thus liquid at a pressure of 1 ton per sq. in. from the force-pump acting on the $2\frac{1}{2}$-inches-diameter piston of the hydraulic press should generate a pressure of 25 tons per sq. in. in the high-pressure cylinder. Actually, on account of friction, the pressure is somewhat less than this.

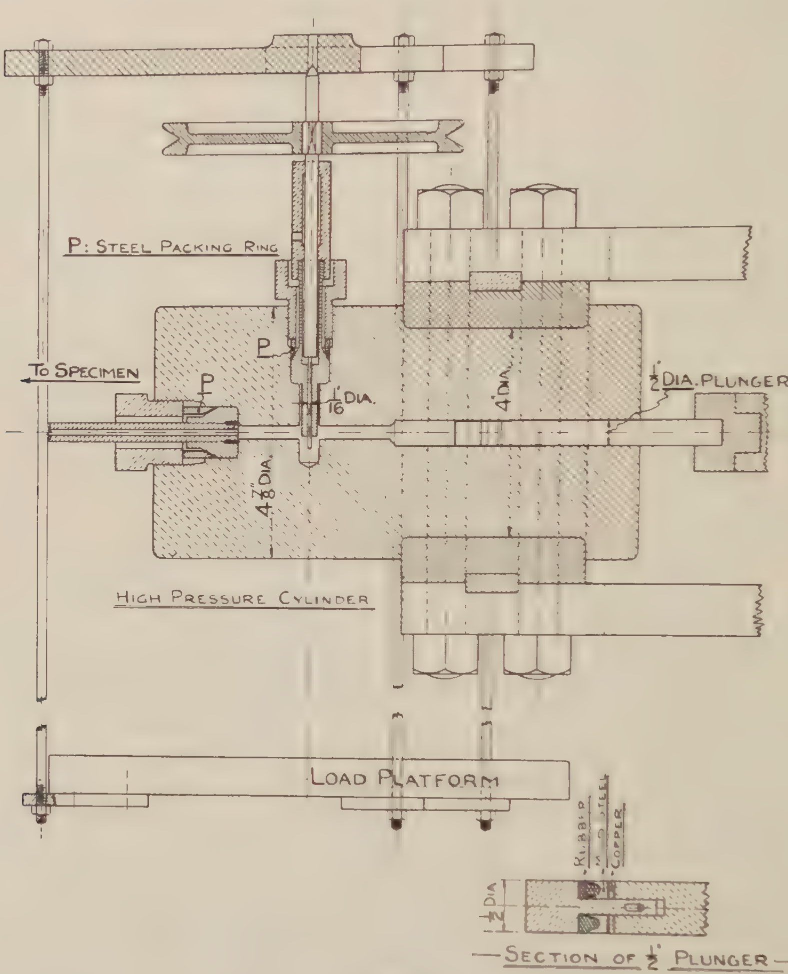


Fig. 4.

The pressure generated in the high-pressure cylinder is measured directly by means of a loaded plunger of diameter 0·0638 in. marked "$\frac{1}{16}$" Dia." in Fig. 4; and the load is applied to this little plunger by adding weights to a platform which is suspended, in the manner shown, from the top of the spindle carrying the little plunger. To eliminate friction this spindle is rotated by means of an electric motor.

Considerable care was given to the construction of the little plunger and of the fitting which carries it. The plunger was made from a knitting-needle, which was "lapped down" parallel and to the correct size. The hole in which it works was

"lapped out" so that the plunger was of a good fit. Afterwards, when the apparatus had been given a trial under a high pressure, the little plunger was removed, and its diameter within about $\frac{1}{2}$ in. from its connection to the load spindle was reduced about 0·0002 in. in order to remove the marks of excessive rubbing which it had acquired while under pressure.

The diameter of the little plunger was obtained accurately to within 0·00005 in. by means of a Newall measuring machine.

The liquid employed in the high-pressure cylinder for communicating the pressure to the specimen cylindrical shell was a very viscous mixture of glucose and glycerine. The liquid employed in the hydraulic press was a mixture of glycerine and water.

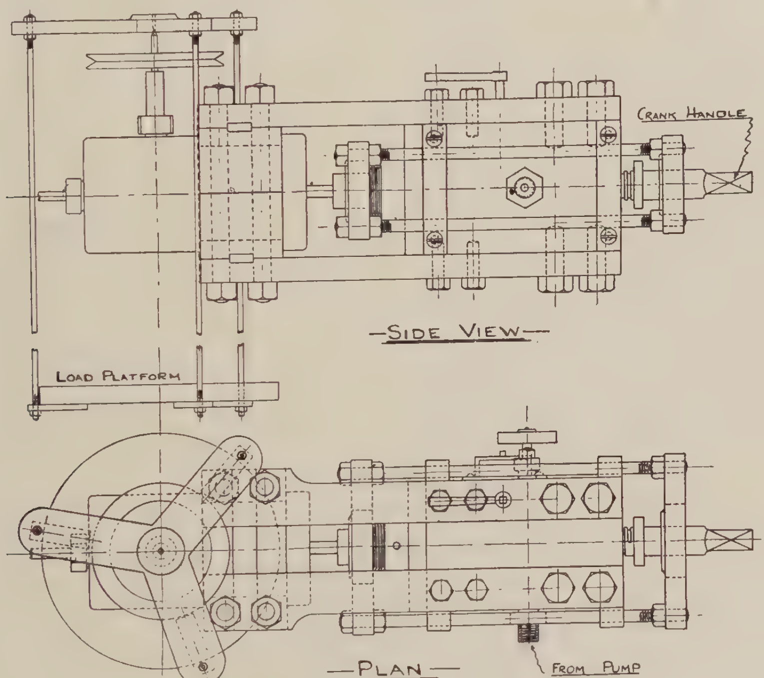


Fig. 5.

The method of packing the $2\frac{1}{2}$-inches-diameter piston, and the $\frac{1}{2}$-inch-diameter plunger, also the methods adopted for making the various joints in the apparatus are clearly shown in the figures.

All the parts of the apparatus exposed to high pressure were made of special alloy steel of high tensile strength. For example, the high-pressure cylinder (Fig. 4) was made from a piece of Nickel Chrome steel, 5 in. in diameter, which was kindly supplied by Messrs Brown Bayley's Steel Works, Ltd., and this was afterwards given by them the necessary heat-treatment.

The hydraulic press was also fitted with screw gear for reversing the $2\frac{1}{2}$-inches-diameter piston, and with a needle-valve for letting out the liquid displaced by the piston.

§ 6. THE MEASUREMENT OF THE LONGITUDINAL STRAINS

In the measurement of the longitudinal and diametral strains produced by hydrostatic pressure inside the various cylindrical shells each specimen was held in an upright position and was connected by special steel tubing to the high-pressure apparatus. The scale employed in conjunction with the longitudinal extensometer was placed about 450 centimetres away from the instrument. It was the usual practice to subject the specimen to pressure after having set up the extensometer in position, and then to leave the apparatus for about 30 min. in order to give it time to settle down to a steady temperature.

In taking a set of readings the pressure was applied in equal increments, the value of the increment depending on the dimensions of the cylinder. In most cases the pressure increment was 1·395 tons per sq. in., which corresponded to the addition of 10 lb. to the load platform. At the commencement the scale reading was generally produced by the pressure corresponding to the weight of the load platform itself, but in several experiments at the first reading there was an initial load placed on the platform.

In applying the pressure it was found necessary to work the pump very slowly; by the adoption of this precaution the unknown additional pressure created in overcoming the inertia of the load in its rise, and by friction, was reduced to a minimum. It was also found that the scale reading attained a maximum value and then remained steady for a few seconds, during which interval it was convenient to take the reading corresponding to the load on the floating plunger. Care was of course exercised in the application of the pressure in order to safeguard the cylinder from being overstrained; two specimens were lost by failure to do this.

Table 2: Longitudinal extensions of specimens 7 and 11

SPECIMEN 7.
Outside diam. = 0·8005″
Inside diam. = 0·400″
EXTENSOMETER:
$d = 0·06295''$, $L = 446$ cm., $l = 4·5$ cm.
Distance between knife-edges = 2·001″

Pressure tons/inch²	Scale reading No.	Scale reading cm.	Scale differences cm.
1·350	1	11·00	(1) − (6) = 4·47
2·745	2	10·11	(2) − (7) = 4·46
4·140	3	9·21	(3) − (8) = 4·43
5·535	4	8·32	(4) − (9) = 4·42
6·930	5	7·43	(5) − (10) = 4·43
8·325	6	6·53	Mean = 4·442
9·720	7	5·65	Scale difference for
11·115	8	4·78	pressure increment
12·510	9	3·90	of 1·395 tons/inch²
13·905	10	3·00	= 0·888 cm.

SPECIMEN 11.
Outside diam. = 0·899″
Inside diam. = 0·3995″
EXTENSOMETER:
$d = 0·06265''$, $L = 398·4$ cm., $l = 5·0$ cm.
Distance between knife-edges = 2·007″

Pressure tons/inch²	Scale reading No.	Scale reading cm.	Scale differences cm.
1·419	1	11·91	(1) − (5) = 2·45
2·887	2	11·30	(2) − (6) = 2·46
4·355	3	10·68	(3) − (7) = 2·45
5·823	4	10·08	(4) − (8) = 2·45
7·291	5	9·46	Mean = 2·4525
8·759	6	8·84	
10·227	7	8·23	Scale difference for
11·695	8	7·63	pressure increment
			of 1·468 tons/inch²
			= 0·6131 cm.

The readings taken in measuring the longitudinal extensions of a cylinder of the 3·48 per cent. nickel steel and of a cylinder of the 0·28 per cent. carbon steel are given in Table 2, in which is also shown the method of arriving at the scale difference for a given pressure increment.

The readings here recorded are representative of those obtained from experiments on the remaining fifteen specimens. It was always possible by careful handling of the pump to obtain a set of scale readings in which the scale differences were constant within one or two per cent. for each equal pressure increment.

A summary of the readings and the results obtained in experiments carried out on the seventeen hollow cylinders are given in Table 3.

Table 3: Summary of results of longitudinal extensional measurements

Dimensions of specimens				Extensometer		Longitudinal strains				$\frac{E}{1-2\sigma} \div 10^6$	Quality of steel
Outside diameter inch	Inside diameter inch	k^2-1	Length inches	Distance between knife-edges inches	$\frac{10^5 d}{4L+2l}$	Pressure increment tons/inch²	Scale difference cm.	Strain $\times 10^5$	Strain for pressure 1 ton/inch² $\times 10^5$		
0·8710	0·5010	2·023	4·0	2·003	3·906	1·468	1·223	2·386	1·626	68·1	Nickel 3·48 %
0·8695	0·4995	2·031	5·1	2·000	3·927	1·468	1·243	2·441	1·663	66·3	
0·8735	0·5010	2·042	8·1	2·003	3·905	1·468	1·246	2·430	1·655	66·3	
0·8743	0·5000	2·056	11·7	2·009	3·875	1·468	1·245	2·402	1·636	66·7	
0·6500	0·4995	0·693	5·1	1·998	3·511	0·419	1·150	2·021	4·823	67·0	
0·7505	0·5000	1·253	8·1	2·004	3·509	0·698	1·059	1·855	2·657	67·3	
0·8005	0·4000	3·004	8·1	2·001	3·510	1·395	0·888	1·558	1·116	66·8	
0·7494	0·5005	1·243	4·0	1·997	3·511	0·698	1·030	1·810	2·592	69·5	Carbon 0·28 %
0·8755	0·5010	2·054	4·0	1·998	3·509	1·395	1·236	2·171	1·556	70·1	
0·7496	0·3760	2·976	4·0	2·000	3·509	1·395	0·860	1·509	1·081	69·6	
0·8990	0·3995	4·065	11·6	2·007	3·905	1·468	0·613	1·193	0·813	67·8	
0·7502	0·5006	1·247	8·0	2·007	3·511	0·698	1·048	1·833	2·626	68·4	Carbon 0·43 %
0·8735	0·5000	2·052	11·6	2·005	3·872	1·468	1·220	2·356	1·605	68·0	
0·8000	0·4000	3·000	8·15	1·995	3·506	1·395	0·875	1·541	1·104	67·6	
0·9005	0·4000	4·065	11·6	2·000	3·911	1·468	0·610	1·192	0·813	67·8	Carbon 0·09 %
0·9363	0·3750	5·234	8·05	2·004	3·509	1·395	0·501	0·877	0·629	68·1	
0·7503	0·2495	8·012	8·05	2·003	3·509	2·093	0·492	0·862	0·412	67·8	

If the values of the strain produced by a pressure of one ton per sq. in., in the eleventh column, are plotted against corresponding values of the reciprocal of $(k^2 - 1)$, in the fourth column, for specimens 2 to 7 of the 3·48 per cent. nickel steel, and for specimens 11 to 17 of the carbon steels, it is found that each of the two series of points thus plotted lies on a straight line passing through the origin; this shows that the longitudinal strain produced by a given internal pressure applied to various cylinders of the same steel depends inversely on the value of $(k^2 - 1)$ for the shell.

The figures in the twelfth column of Table 3, under the heading $E/(1 - 2\sigma)$, were obtained by dividing 2240 by the value of $(k^2 - 1)$ for the specimen, and by the strain produced by a pressure of one ton per sq. in. It is to be noted how closely

the figures thus calculated agree amongst themselves for specimens 2 to 7, and for specimens 11 to 17, the figures being within 1 per cent. of the mean values $66\cdot7 \times 10^6$ and $67\cdot9 \times 10^6$ respectively.

The figures in italics in the twelfth column of Table 3 were obtained from readings taken in experimenting on specimens 1, 8, 9 and 10, all of which were 4 in. long between the screwed ends; and these figures are about 2 per cent. greater than the corresponding values already given.

This question of the effect of the ends of the cylinder on the extension was investigated very carefully in the case of the nickel steel for specimens 1 to 4, all of which had practically the same $(k^2 - 1)$, the length of each taken in order being 4, 5·1, 8·1 and 11·7 in. respectively. It would therefore appear from the values given in the twelfth column of Table 3 for these specimens that the ends of the specimen do not influence the longitudinal extensions to any great extent if its length is greater than 5 in. But most of the specimens dealt with were longer than this; in Table 3 it is seen that four were about 11 in. and seven about 8 in. long, the lengths being measured between the screwed ends of the specimens.

As a general rule the agreement between the various sets of readings for a given specimen was very good, but in two or three cases it was observed that in taking a second set of readings immediately after the first set the scale differences were about 2 per cent. greater, though other subsequent sets of readings appeared to agree with the first set within 1 per cent. Having noted this, the author was careful to repeat each experiment several times before removing the specimen from the apparatus.

It should be mentioned here that after the apparatus had been in use for about six months the high-pressure cylinder developed a leak, the result of a flaw in the material, and this necessitated replacing the cylinder by a new one. A new floating plunger was also made together with the fitting in which it works. This explains why in Table 3 the pressure increments are in some cases 1·468 tons per sq. in. and in others 1·395 tons per sq. in. In the former, the pressure is due to the application of a load of 10 lb. on the first plunger of diameter 0·06225 in., while in the latter, the pressure is due to the application of 10 lb. to the second of diameter 0·0638 in.

§ 7. THE MEASUREMENT OF THE DIAMETRAL STRAINS

The measurement of the diametral extensions has been done by other investigators with different objects in view. Professors Cook and Robertson carried out in 1910 experiments on "The Strength of Thick Hollow Cylinders under Internal Pressure"*, and Prof. P. W. Bridgman published in 1920 an account of "An Experiment in One-piece Gun Construction."† In both these investigations the measurement of the diametral strains was used chiefly as a means of detecting the pressure at which the material of the thick hollow cylinder yielded. An experiment described by Cook and Robertson in the above paper throws some light on the

* *Engineering*, **92**, 786 (1911).

† *Mining and Metallurgy*, No. 158, Section 14 (1920).

shape of a specimen under internal pressure; it showed that the diametral extensions at various sections are constant near the central portion of the specimen, but near the enlarged ends the extensions are slightly greater. There would thus appear to be some justification for assuming in the case of a long cylindrical shell that each generating line remains straight after the application of the pressure.

The experiments recorded in this section of the present paper were carried out on much the same lines as those already described in the previous section. The specimen was again held in an upright position while under test. The lateral extensometer was placed in contact with the specimen at or near to its middle section, and the instrument was supported lightly by three thin pieces of cord made fast to a plate at the top.

The horizontal scale employed in conjunction with the extensometer was placed about 300 cm. away from the instrument. Since the extensions were much larger in these experiments, it was not necessary to place the scale so far away as in the measurement of the longitudinal extensions.

Table 4: Diametral extensions of specimens 3 and 12

Specimen 3. Outside diam. = 0·8735″ Inside diam. = 0·5010″ Extensometer: d = 0·02584″, L = 279·5 cm., l = 1·9 cm.				Specimen 12. Outside diam. = 0·7502″ Inside diam. = 0·5006″ Extensometer: d = 0·02584″, L = 327·2 cm., l = 1·9 cm.			
Pressure tons/inch²	Scale reading No.	Scale reading cm.	Scale differences cm.	Pressure tons/inch²	Scale reading No.	Scale reading cm.	Scale differences cm.
1·350	1	28·11	(1) − (5) = 13·83	2·048	1	24·84	(1) − (6) = 8·26
2·745	2	24·70	(2) − (6) = 13·86	2·467	2	23·20	(2) − (7) = 8·30
4·140	3	21·21	(3) − (7) = 13·81	2·886	3	21·57	(3) − (8) = 8·29
5·535	4	17·80	(4) − (8) = 13·85	3·305	4	19·92	(4) − (9) = 8·37
6·930	5	14·28	Mean = 13·8375	3·724	5	18·26	(5) − (10) = 8·37
8·325	6	10·84		4·143	6	16·58	Mean = 8·318
9·720	7	7·40	Scale difference for pressure increment of 1·395 tons/inch² = 3·459 cm.	4·562	7	14·90	Scale difference for pressure increment of 0·419 ton/inch² = 1·664 cm.
11·115	8	3·95		4·981	8	13·28	
				5·400	9	11·55	
				5·819	10	9·89	

The readings taken in measuring the diametral extensions of a cylinder of the 3·48 per cent. nickel steel and of a cylinder of the 0·43 per cent. carbon steel are given in Table 4, in which the method of arriving at the scale difference for a given pressure increment is again shown. The readings here recorded are representative of those obtained from experiments on the remaining thirteen specimens. A summary of the readings and the results obtained in experiments carried out on the fifteen specimens is given in Table 5.

If, as in the previous section, the strains produced by a pressure of one ton per sq. in. are plotted against corresponding values of the reciprocal of $(k^2 - 1)$ for the specimens 2 to 7, and for 11 to 17 respectively, it is found that each of the series of

points lies on a straight line passing through the origin; this shows that the diametral strain produced by a given pressure applied to various cylinders of the same steel depends inversely on the value of $(k^2 - 1)$ for the shell.

The figures given in the eleventh column of Table 5, under the heading $E/(2 - \sigma)$, were obtained by dividing 2240 by the value of $(k^2 - 1)$ for the specimen and by the diametral strain produced by a pressure of one ton per sq. in. It is to be noted how closely the figures thus calculated agree amongst themselves for specimens 2 to 7 and for specimens 11 to 17. These values are within 1 per cent. of the mean values $16{\cdot}81 \times 10^6$ and $17{\cdot}22 \times 10^6$ respectively.

Table 5: Summary of results of diametral extensional measurements

Specimen	Dimensions of specimens: Outside diameter inch	Inside diameter inch	$k^2 - 1$	Length inches	Extensometer $\frac{10^5 d}{4L - 2l}$	Diametral strains: Pressure increment tons inch⁻	Scale difference cm.	Diametral extension 10 inch	Strain for pressure 1 ton inch × 10	$\frac{E}{2 - \sigma}$ 10	e_2/e_3	Qu… of …
1	0·8710	0·5010	2·023	4·0	2·305	1·395	3·425	7·894	6·497	*17·04*	4·00	
2	0·8695	0·4995	2·031	5·1	2·296	1·395	3·461	7·945	6·549	16·84	3·94	
3	0·8735	0·5010	2·042	8·1	2·305	1·395	3·459	7·971	6·542	16·77	3·95	Ni…
5	0·6500	0·4995	0·693	5·1	2·014	0·419	2·605	5·246	19·26	16·78	3·99	3·4…
6	0·7505	0·5000	1·253	8·1	1·969	0·419	1·695	3·338	10·61	16·84	3·99	
7	0·8005	0·4000	3·004	8·1	2·230	1·395	2·223	4·958	4·439	16·80	3·98	
8	0·7494	0·5005	1·243	4·0	2·039	0·419	1·582	3·226	10·28	*17·54*	3·97	
9	0·8755	0·5010	2·054	4·0	2·251	1·395	3·380	7·607	6·229	*17·51*	4·00	Ca…
10	0·7496	0·3760	2·976	4·0	2·255	1·395	2·010	4·532	4·333	*17·37*	4·01	0·2…
11	0·8990	0·3995	4·065	11·6	2·039	1·395	1·981	4·040	3·221	17·11	3·96	
12	0·7502	0·5006	1·247	8·0	1·969	0·419	1·664	3·275	10·42	17·24	3·97	Ca…
14	0·8000	0·4000	3·000	8·15	2·230	1·395	2·170	4·840	4·336	17·22	3·93	0·4…
15	0·9005	0·4000	4·065	11·6	2·041	1·395	1·977	4·034	3·212	17·16	3·95	Ca…
16	0·9363	0·3750	5·234	8·05	1·914	1·395	1·688	3·230	2·474	17·30	3·93	0·0…
17	0·7503	0·2495	8·012	8·05	1·730	1·395	0·980	1·694	1·619	17·27	3·93	

The figures in the eleventh column for specimens 1, 8, 9 and 10 are slightly greater than the corresponding mean values just given. But these specimens are the short ones. The ends of the cylinder appear to have only a slight effect on the extensional measurements at the middle section for cylinders 4 in. in length between the screwed ends.

The ratio of the diametral strain e_2 to the longitudinal strain e_3, both strains being produced by the same internal pressure, is given in the twelfth column of Table 5 for each specimen. Leaving out of consideration the short specimens, it is here seen that this ratio is constant and equal to 3·96 within 1 per cent. for the four varieties of steel dealt with, a result which agrees with equation (8) on page 367.

Let us now see how the result of the investigation fits the usually accepted theory. Writing $(2 - \sigma)/(1 - 2\sigma) = 3{\cdot}96$, we find that $\sigma = 0{\cdot}283$ for the four varieties

of steel. Substituting this in equations (5) and (7), we find that in the case of the nickel steel, since $E/(1-2\sigma) = 66{\cdot}7 \times 10^6$, $E = 28{\cdot}95 \times 10^6$ lb. per sq. in., and since $E/(2-\sigma) = 16{\cdot}81 \times 10^6$, $E = 28{\cdot}87 \times 10^6$ lb. per sq. in.; also, in the case of carbon steels, since $E/(1-2\sigma) = 67{\cdot}9 \times 10^6$, $E = 29{\cdot}47 \times 10^6$ lb. per sq. in., and since $E/(2-\sigma) = 17{\cdot}22 \times 10^6$, $E = 29{\cdot}56 \times 10^6$ lb. per sq. in.; and these values of Young's modulus agree fairly well with the usually accepted values for similar steels.

The actual amount of the diametral extension of a cylinder under pressure as given in the ninth column of Table 5 is interesting. In the case of specimen 3, with outside diameter 0·8735 in. a pressure of 13·95 tons per sq. in. increases the diameter by an amount equal to 0·0007971 in., which is less than 0·1 per cent. of the original diameter. Also, in case of specimen 12, with outside diameter 0·7502 in., a pressure of 8·38 tons per sq. in. increases the diameter by an amount equal to 0·000655 in., an amount which is also less than 0·1 per cent. of the original diameter. If the inside diametral extensions of the specimens are of the same order of magnitude there would appear to be some justification for the assumption made in deriving the theory that the alterations in the radial dimensions of the shell are inappreciable after the application of the pressure. The corresponding values of the inside diametral extensions of each of the above cylinders are 0·001145 in. and 0·00085 in. respectively, these being calculated on the assumption that the present theory is correct.

§ 8. E AND σ BY DIRECT TENSION EXPERIMENT

An interesting comparison may be made between these values of E and σ and those obtained by direct tension of several of the same thick cylinders in a single-lever testing machine employing the same extensometers.

The testing machine could apply loads up to 2 tons, and also loads up to 10 tons could be applied by means of two jockey-weights, the one of weight 168 lb. and the other 672 lb., the two together giving the greater range of load, the smaller alone the lesser range; the beam had two corresponding load scales. The load was applied by a screw operated by worm gearing at the front of the machine. The specimen was secured to the loading gear, and to the beam, by means of an axial loading attachment at both ends, this being held in the grips as shown in Fig. 1.

Before commencing this part of the work a thorough examination of the machine was made. Both jockey-weights were weighed separately, and the weight of each was found to be correct. Then the loading gear was unshipped, and weights up to 0·705 ton were hung from the load shackle of the beam; and in balancing the beam by the smaller jockey-weight it was found that its load scale was correct within the range of load suspended; it was therefore inferred that both the load scales were correct.

The two extensometers were employed in exactly the same manner as has already been described in the previous sections, and they were shielded from the rays of the sun by a large screen.

The effect of the oscillation of the beam of the testing machine on the working of the extensometers was tested and was found to be inappreciable provided the oscillations were not large. Also the effect of a slight tremor, which was always present when the load was being changed, was tested; and this was found to be very slight indeed provided the change of the load was done with care.

The following are the results of experiments carried out on eight of the thick hollow cylinders:

Table 6: Results of tension experiments

Specimen	3	6	7	11	12	14	15	16
$E \div 10^6$	29·8	29·7	29·5	30·5	30·5	30·5	30·4	30·0
Poisson's ratio	0·292	0·293	0·291	0·294	0·291	0·291	0·294	0·290

from which it would appear that for the 3·48 per cent. nickel steel $E = 29{\cdot}7 \times 10^6$ lb. per sq. in., and for the carbon steels $E = 30{\cdot}4 \times 10^6$ lb. per sq. in., and Poisson's ratio is 0·292 for the four steels. Comparing these with the results of the pressure experiments it is found that

$$(E \text{ by testing machine})/(E \text{ by pressure experiments}) = 1{\cdot}027,$$

and

$$(\sigma \text{ by testing machine})/(\sigma \text{ by pressure experiments}) = 1{\cdot}032.$$

In view of these discrepancies it was thought advisable to verify the accuracy of the two extensometers.

§ 9. VERIFICATION OF THE ACCURACY OF THE EXTENSOMETERS

The lateral extensometer had already been checked by Prof. Lamb who mounted the instrument in a measuring machine, and the results of his experiments showed that the extensometer was capable of making a measurement of a lateral extension of 0·0002 in. to within 0·3 per cent.*

The performance of the 2-in. extensometer was tested against that of three other instruments.

1. Determinations of Young's modulus of four different steel bars by both the 2-in. extensometer and a 4-in. roller extensometer with rollers 0·125 in. diameter were found to be in agreement to within $\frac{1}{2}$ per cent.

2. Determinations of Young's modulus of a bar of the 3·48 per cent. nickel steel by both the 2-in. instrument and a Ewing's extensometer gave the values $29{\cdot}7 \times 10^6$ and $29{\cdot}4 \times 10^6$ lb. per sq. in. respectively. But it should be pointed out that the scale difference for a load increment of $\frac{1}{2}$ ton was about 0·48 in Ewing's extensometer while it was 3·65 in the roller extensometer, the latter being capable of a much finer degree of accuracy.

3. Employing the same bar of nickel steel, Young's modulus was again determined by a 4-in. Marten's extensometer. This gave E as $29{\cdot}4 \times 10^6$ lb. per sq. in.; and in this case for a load increment of $\frac{1}{2}$ ton the scale difference was about 0·90. For the reason that the exact dimension of the distance across the double knife-

* *Engineering*, **119**, 207 (1925).

edges, when in contact with the surface of the specimen, must always be slightly less than the measured distance, Young's modulus might easily be $\frac{1}{2}$ per cent. less than the value found by the roller extensometer.

In consideration of these experiments and of the much greater degree of sensitiveness of the roller extensometer compared with either Ewing's or Marten's instruments, one is perhaps justified in regarding the longitudinal extensional measurements of the pressure experiments as being substantially accurate.

If then it is granted that as far as possible experimental errors have been eliminated, and that the theory of the thick cylindrical shell under pressure is correct, the discrepancies in the values of Young's modulus and Poisson's ratio determined by the testing machine and pressure experiments are perhaps due to the non-isotropic nature of the material.

§ 10. CONCLUSIONS

The constants under the heading $E/(1-2\sigma)$ in Table 3 and under the heading $E/(2-\sigma)$ in Table 5 are evidently independent of the thickness of the shell, and it may therefore be concluded that they would hold for a thin shell.

On the assumption that the material is non-isotropic, and employing suffixes (2) and (3) for lateral and longitudinal directions respectively, one can write for a thin shell of the nickel steel:

$$1/E_3 - 2\sigma_2/E_2 = 1/66{\cdot}7 \times 10^6,$$

and

$$2/E_2 - \sigma_3/E_3 = 1/16{\cdot}81 \times 10^6,$$

from which, making use of the data $E_3 = 29{\cdot}7 \times 10^6$ and $\sigma_3 = 0{\cdot}292$ as found in the testing machine, we calculate E_2 to be $28{\cdot}9 \times 10^6$, and σ_2 to be $0{\cdot}270$. Also, for the three carbon steels, in a similar manner, we calculate E_2 to be $29{\cdot}55 \times 10^6$, and σ_2 to be $0{\cdot}269$. Whence the following ratios:

$$E_2/E_3 = 0{\cdot}973, \text{ and } \sigma_2/\sigma_3 = 0{\cdot}921.$$

It has been suggested* that the bulk modulus K might be obtained from the longitudinal strains of a thick cylindrical shell under internal pressure, K being equal to $E/3\,(1-2\sigma)$.

From the twelfth column of Table 3, dividing by 3, K for the nickel steel is $22{\cdot}2 \times 10^6$ lb. per sq. in., and for the carbon steels $22{\cdot}6 \times 10^6$ lb. per sq. in. Substituting in the formula for K the values of E and σ which were obtained in the direct tension experiments on these steels, we obtain that K is equal to $23{\cdot}8 \times 10^6$ and $24{\cdot}3 \times 10^6$ respectively, values which are 7 per cent. greater than those already given. But this discrepancy may in a measure be accounted for by slight inaccuracies in the measurements of the longitudinal and lateral strains in determining E and σ in the testing machine; if, for example, the measured longitudinal strains are $\frac{1}{2}$ per cent. less, and the lateral strains are 1 per cent. more than the correct amounts, the value of K as calculated by means of the above formula is about $2\frac{1}{2}$ per cent. too great. Allowing for this, it would still appear that the

* A. Mallock, *Trans. Roy. Soc.* (1904); H. Lamb, *Statics*, p. 333.

determination of the bulk modulus by the pressure method yields a low value for this elastic constant.

From the investigations it would appear that when the usually accepted theory is applied in determining the elastic constants of a material, such as steel, by subjecting a thick hollow cylinder to internal pressure, the method, apart from considerations of the difficulty of such experiments, does not give values which are in perfect agreement with those obtained more conveniently in a tensile testing machine.

§ 11. ACKNOWLEDGMENT

The best thanks of the author are due to the Council of the East London College and to Prof. E. H. Lamb for the facilities which have been placed at his disposal for carrying out the work described in this paper.

DISCUSSION

Mr J. P. ANDREWS: Mr Wedgwood has attributed the difference between the elastic constants obtained by his pressure method and by the tensile testing machine to anistropy in his specimens. Since this appears to be of a similar character in all cases, it would probably be due to one or more of the processes required to produce the cylinders employed. Could Mr Wedgwood tell us sufficient about these processes to enable us to gather information as to the effect of working, etc. on the properties of the material? I would like to remark that I have used a lateral extensometer of Prof. Lamb's design, similar to that used by Mr Wedgwood. In my opinion it is amply capable of the accurate measurement of small extensions such as those met with in this investigation. What precautions did Mr Wedgwood take to eliminate the effect of temperature upon this instrument?

Mr SEARS: Mr Wedgwood has described a very good and a very difficult piece of work. The only point which occurred to me to criticise in listening to his paper related to the accuracy of measurement of the pressures employed. I understood him to say that an accuracy of 0·1 per cent. was obtainable, and that he used a plunger 1/16th in. in diameter. To obtain this accuracy of pressure measurement would mean knowing not only the diameter of the plunger, but the mean diameter of the plunger and the hole in which it works, to an accuracy of 1/32,000th in. Even if the plunger were measured to this accuracy it would be very difficult to do the same for the hole. The discrepancy between his values for E and σ found by pressure measurements in the tensile testing machine is of course much greater than the above amount, but it is possible that the accuracy of pressure measurement might have some bearing on the result.

Dr FERGUSON: Mr Wedgwood's very careful and ingenious experiments show clearly that there is an outstanding discrepancy between the values of Young's modulus and of Poisson's ratio as determined by these two methods. Is it not possible that, under these high pressures, there will be a swelling of the cylinder into a barrel form which is not contemplated by the elementary theory?

Mr APPLEYARD: Are the observed differences in E and σ to be attributed to any measurable extent to differences in the ratio of length to diameter of the specimen tubes?

AUTHOR'S reply: In reply to Mr Andrews: I hardly think that the various processes of the manufacture of the specimens had any effect on the elastic properties of the material of the shells. Had this been the case the discrepancies mentioned would have been greater in the thinner shells. The work of drilling the holes in the longer shells was done very well by the Birmingham Small Arms Company. I consider that the precautions described in the paper were sufficient to eliminate the effects of temperature upon the working of the extensometers.

In reply to Mr Sears: Prof. P. W. Bridgman, who developed the device for measuring the pressures, claimed for it an accuracy of 0·1 per cent. He obtained the "effective area" of the little plunger of his high pressure apparatus hydrostatically by balancing the plunger, which was suitably loaded, against another and larger floating plunger of known diameter. In the present work the diameter of the little plunger was obtained directly by means of a Newall measuring machine, which was capable of making a measurement to the nearest 2×10^{-5} in. The diameter obtained in this way should be correct to within one in three thousand; and if the plunger is a good fit in the hole the measurement of the diameter should be sufficiently accurate for the purpose of this work. The plunger, if one can judge by surface marks, had every appearance of being a good fit when the apparatus was under pressure.

In reply to Dr FERGUSON: Experiments made by Professors Cook and Robertson in 1910 on thick hollow cylinders under internal pressure showed that the diametral extensions are constant near the central portion of the specimen, but near the corners at the enlarged ends the extensions are slightly greater. There would thus appear to be some justification for assuming in developing the theory that each generating line of the cylinder remains straight after the application of the pressure.

In reply to Mr Appleyard: The question of the effects of the ends of the cylinder on the extensions, both longitudinal and diametral, was investigated and I found that in specimens in which the length was about six times the outside diameter the ends did not influence the measured extensions.

A SIMPLE METHOD OF FITTING A STRAIGHT LINE TO A SERIES OF OBSERVATIONS

By J. H. AWBERY, B.A., B.Sc.
Physics Department, The National Physical Laboratory

Received February 12, 1929. *Read and discussed May* 10, 1929

ABSTRACT. A method is described for determining the constants in a straight line to represent a series of data. It has a rational basis, and can be carried out in one-third or one-quarter of the time required to apply the method of least squares; it avoids the multiplications and squarings essential to that method, and has the advantage that at no stage does it necessitate the use of more significant figures than are presented in the data to be analysed. The results of the method are compared with those obtained by the method of least squares, on some accurate, and also on some very scattered, observational material. In both cases, the saving of time and labour lead to a negligibly small decrease in accuracy, as compared with the older method.

The determination of the best linear equation to fit a set of observations is a relatively frequent requirement, since it is involved in all problems of determining the constants in a two-constant equation. For example, the equations $y = ax^m$, $y = ae^{bx}$ or $y = ax/(x^2 + b^2)$ reduce to linear equations when written in the forms $\log y = \log a + m \log x$, $\log y = \log a + bx$, and $x/y = x^2/a + b^2/a$, respectively.

The usual method of dealing with such cases, when the work cannot be conveniently carried out graphically, is to apply the method of least squares, and the object of the present note is to suggest a method which is far less laborious and appears to give results nearly identical with those of the least square method.

In the first place, we may find one point on the line by taking the "centre of gravity" of all the points. That is, if the observations are $y = y_1, y_2, y_3 \ldots y_n$ when x has the values $x_1, x_2, x_3 \ldots x_n$, then a point on the line is

$$\bar{y} = \sum_1^n y_r/n, \quad \bar{x} = \sum_1^n x_r/n,$$

where n is the number of observations. We may note that this point would also be on the line given by the method of least squares, for, taking

$$y = mx + c,$$

as the standard equation to the line, the coefficient of c in each observational equation is unity, so that the first "normal equation" is

$$\sum_1^n y_r = m \sum_1^n x_r + nc,$$

and the point determined above is clearly consistent with this equation.

We note that the coordinates of this "base-point" need only be determined to the same number of significant figures as the observations themselves.

The next step is to transfer the origin to the base point, by subtracting $\bar{y}$ and $\bar{x}$ from each observation, and this reduces the problem to the determination of m in the equation $y = mx$. One method that might suggest itself would be to divide each y_r in this simple problem by the corresponding x_r, so as to obtain n estimates of m, and to average these; clearly this would not be fair, since the estimate of m formed as the ratio of two small quantities is entitled to less weight than the estimate from a large x_r and y_r. If we weight the estimates according to their values of x_r, we reproduce the value of y_r, and thus obtain the very simple rule: To obtain m, add up all the values of x_r, and all the values of y_r, and divide the latter sum by the former.

As regards simplicity, this procedure contrasts strikingly with the method of least squares, which would solve the problem by multiplying every observational equation by x_r and forming the normal equation

$$\Sigma x_r y_r = m \Sigma x_r^2.$$

The formation of $2n$ multi-figured products and the addition of the two columns are replaced by the addition of two columns, each consisting of numbers with a smaller number of significant figures.

It is clear that the result cannot differ much from that furnished by the least square method, for in the latter, m is given as $\Sigma x_r y_r / \Sigma x_r^2$. But y_r is approximately equal to mx_r, so that the method of least squares gives approximately

$$m = \Sigma m x_r^2 / \Sigma x_r^2,$$

whereas we propose taking

$$m = \Sigma m x_r / \Sigma x_r.$$

When the origin is removed to the centre of gravity of the observations, it is clear that some figures will become negative, others positive. If a line which lies in the first and third quadrants be considered, the estimation of m for a point in the third quadrant is unaffected by changing the signs of both x_r and y_r. If, however, the point is in the second or fourth quadrants, the point tends to rotate the line anti-clockwise, so that only one sign must be changed.

We may sum up the whole procedure as follows, commencing with the table of values of x_r and y_r. (1) Sum both columns algebraically, and take the mean of each, viz. $\bar{x}$ and $\bar{y}$. (2) Subtract $\bar{x}$ from each x_r and $\bar{y}$ from each y_r, to give x_r' and y_r'. (3) Make every x_r' positive, at the same time changing the sign of y_r' when x_r' is negative. (4)* Sum the resulting two columns algebraically, and divide one total by the other. The quotient is the value of m in the equation

$$y = mx + c,$$

while c is determined since the point $(\bar{x}, \bar{y})$ is on the line. At no stage is it necessary to use more significant figures than are used to express the observations, this state of affairs being in marked contrast to the necessities of the least square method.

* Since the deviations are measured from the centre of gravity, it is clear that the arithmetic could be reduced by using only half the observational material, viz. those for which x_r' is positive originally.

The method has been tested in several cases, both where the observations are of high accuracy, and also where they are so scattered that the selection of the best line to represent them is a matter of some difficulty. The latter case arises sometimes when a small correction is to be determined by internal analysis of a series of observations. Each pair of observations gives an estimate of the correction, and the estimates may differ by very large fractions of the correction itself, although the divergences are small compared with the quantity to be corrected; nevertheless, it will be desirable to smooth the corrections, and apply them to the experiments separately.

Example I. The quantities t and R in Table 1 are the measured values of the resistance R of a coil at the temperatures t, and it was desired to calculate the constants in the line $R = at + b$.

Table 1: Comparison with results obtained by method of least squares

t	R (observed)	R calc.		Residuals	
		Least squares	New method	Least squares	New method
14·161	48·9169	48·9178	48·9180	− 9	− 11
14·211	48·9265	48·9269	48·9270	− 4	− 5
14·633	49·0039	49·0035	49·0037	+ 4	+ 2
14·642	49·0055	49·0052	49·0053	+ 3	+ 2
15·825	49·2205	49·2200	49·2201	+ 5	+ 4
15·833	49·2218	49·2215	49·2216	+ 3	+ 2
19·289	49·8496	49·8492	49·8492	+ 4	+ 4
19·292	49·8499	49·8497	49·8497	+ 2	+ 2
19·542	49·8954	49·8951	49·8951	+ 3	+ 3
19·574	49·9014	49·9009	49·9010	+ 5	+ 4
19·612	49·9083	49·9078	49·9079	+ 5	+ 4
20·414	50·0532	50·0535	50·0535	− 3	− 3
20·431	50·0566	50·0566	50·0566	± 0	± 0
20·449	50·0596	50·0598	50·0599	− 2	− 3
20·670	50·0997	50·1000	50·1000	− 3	− 3
20·688	50·1029	50·1033	50·1033	− 4	− 4
22·036	50·3483	50·3481	50·3481	+ 2	+ 2
22·062	50·3530	50·3528	50·3528	+ 2	+ 2
24·024	50·7093	50·7091	50·7091	+ 2	+ 2
24·029	50·7102	50·7100	50·7100	+ 2	+ 2
24·060	50·7155	50·7157	50·7156	− 2	− 1
24·079	50·7187	50·7191	50·7191	− 4	− 4
26·136	51·0928	51·0927	51·0926	+ 1	+ 2
26·159	51·0969	51·0969	51·0968	± 0	+ 1
27·299	51·3041	51·3039	51·3038	+ 2	+ 3
27·300	51·3042	51·3041	51·3040	+ 1	+ 2

The calculation by the method of least squares occupied some three hours, and gave the result $R = 46{\cdot}3459\,(1 + 0{\cdot}391879\,t)$. The present method was also applied. It occupied about one hour, and gave the result $R = 46{\cdot}3463\,(1 + 0{\cdot}391835\,t)$. The full calculation by the present method is shown in Table 2, whilst the real degree of agreement between the two equations may perhaps be better estimated from the last four columns of Table 1, which give successively the calculated values of R from the two equations, and the residuals (R_{observed} less $R_{\text{calculated}}$), the unit for the residuals being one unit in the fourth decimal place.

Table 2: Example of method of calculation

t	R	$(t - 20\cdot633)$	$\pm (R - 50\cdot0933)$
14·161	48·9169	6·472	1·1764
14·211	48·9265	6·422	1·1668
14·633	49·0039	6·000	1·0894
14·642	49·0055	5·991	1·0878
15·825	49·2205	4·808	0·8728
15·833	49·2218	4·800	0·8715
19·289	49·8496	1·344	0·2437
19·292	49·8499	1·341	0·2434
19·542	49·8954	1·091	0·1979
19·574	49·9014	1·059	0·1919
19·612	49·9083	1·021	0·1850
20·414	50·0532	0·219	0·0401
20·431	50·0566	0·202	0·0367
20·449	50·0596	0·184	0·0337
20·670	50·0997	0·037	0·0064
20·688	50·1029	0·055	0·0096
22·036	50·3483	1·403	0·2550
22·062	50·3530	1·429	0·2597
24·024	50·7093	3·391	0·6160
24·029	50·7102	3·396	0·6169
24·060	50·7155	3·427	0·6222
24·079	50·7187	3·446	0·6254
26·136	51·0928	5·503	0·9995
26·159	51·0969	5·526	1·0036
27·299	51·3041	6·666	1·2108
27·300	51·3042	6·667	1·2109
Mean t = 20·633	Mean R = 50·0933	Sum = 81·900	Sum = 14·8731

$$\frac{14\cdot8731}{81\cdot900} - 0\cdot181601.$$

When

$$t = 0,\ R = 50\cdot0933 - (0\cdot181601 \times 20\cdot633) = 46\cdot3463.$$

Hence linear relation is

$$R = 46\cdot3463 + 0\cdot181601\, t = 46\cdot3463\ (1 + 0\cdot391835\, t).$$

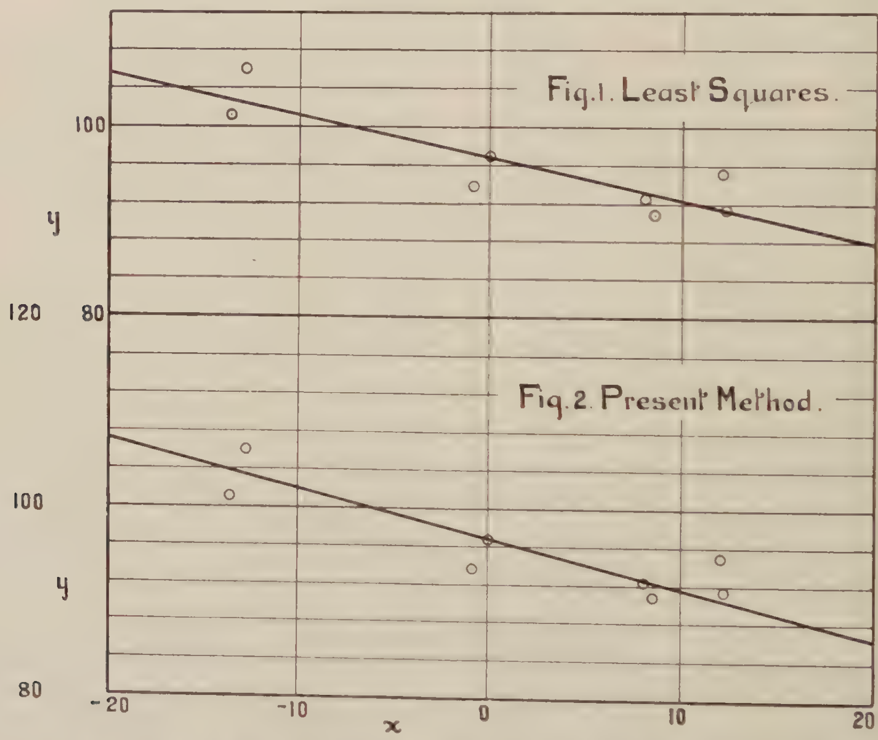
In order to find whether the satisfactory agreement with the least square method was fortuitous in this case, an exactly similar comparison was made with two other series of experimental results. They were actually similar measurements of another resistance coil, and a series of measurements of the two in series. The conclusions may be summarised as follows.

For the measurements quoted above, the mean residual was 0·6, whilst with least squares it was 0·1; meeting the latter method on its chosen ground, the mean square residuals were 12·7 by the new method, and 12·1 by least squares. For the other two series, the present method gave mean residuals of 1·2 and 2·5, as against − 0·5 and − 0·1 by least squares, the mean square residual being 15·6 and 45·7, as against 13·9 and 32·7. These results are quite satisfactory, and in fact the new method gives results, for the present series, which could not really be distinguished from those obtained by least squares. The actual consideration of the constants is also of interest. Extrapolating back to 0° C. the resistances may be compared as follows.

Table 3: Consistency of results

	Resistance		Temperature coefficient of wire	
	Least squares	Present method	Least squares	Present method
1st section	46·4894	46·4897	0·391879	0·391829
2nd section	46·3459	46·3463	0·391879	0·391835
1st section + 2nd section	92·8353	92·8360	—	—
1st section + 2nd section measured together in series 3	92·8359	92·8360	0·391869	0·391835

The individual estimates of R_0 agree with those of the least square method to within 0·0007 ohm = 2/1000° C. approximately, but the sum of the two coils agrees even better with the resistance in series than it does when the method of least squares is used. The coefficients of resistance are in agreement throughout in either method, but the present method gives consistently lower values.



Example II. This is an example of the scattered type of observations. Here very high accuracy is clearly unnecessary, and the saving of time may frequently be considered even more worth while than it is with observations of high precision. The points through which it is desired to draw a straight line are shown in the figure, and are set out in Table 4, which also shows the results calculated from the

ines $y = 96\cdot 7 - 0\cdot 44_4 x$ and $y = 96\cdot 8 - 0\cdot 51_4 x$, of which the former results from he method of least squares and the latter from the present method. Even in this ase, the mean square deviation is only increased from 5·65 to 6·08 by the use of the ess laborious method. (The mean deviations in both methods are 0·1 and − 0·1).

Table 4: Data of Example II

x	y	y (calculated)	
		Least squares	Present method
$-13\cdot 5_5$	101·2	102·7	103·8
$-12\cdot 8$	106·0	102·4	103·4
$-0\cdot 8$	93·7	97·1	97·2
0·0	96·8	96·7	96·8
8·1	92·5	93·1	92·6
8·6	90·8	92·9	92·4
12·1	95·0	91·3	90·6
$12\cdot 2_5$	91·4	91·3	90·5

It may be of interest to compare the lines given by these methods with the lines drawn "by eye" by two observers. One obtained

$$y = 96\cdot 5 - 0\cdot 39_3 x,$$

and another

$$y = 96\cdot 6 - 0\cdot 51_6 x.$$

Thus these lines differ in slope more than do the two lines given by the arithmetical methods.

On the whole, it appears that when the data are of high accuracy, the line given will be extremely close to that obtained by the method of least squares; when they are of low accuracy, the two methods agree sufficiently well to justify the time saved at the expense of the accuracy, and in general it appears that the method proposed may well replace that of least squares for practically all work where the latter method would now be used. The saving in labour is greater for extensive series than for short ones.

DISCUSSION

Dr J. S. G. Thomas said that of the many straight lines that could be drawn through a number of points one could undoubtedly be found by the author's method, but the title of the paper did not indicate what was its peculiarity. Was it the best straight line?

Dr Ferguson: I assume that Mr Awbery's view is that the lines determined by the least squares method are to be taken as standard lines, and he is simply showing that lines determined by his method are unique, in a sense in which lines determined by the fitting of a stretched thread are not unique, and that they are in good agreement with the "least squares" lines, and may be drawn at the expense of considerably less labour.

Mr J. E. Sears: The method of least squares is liable to give a misleading estimate of accuracy unless the number of observations involved is very large, and therefore unless this condition is fulfilled the extra labour involved in applying the method of least squares is hardly justified if a simpler alternative method is available which is known to give sufficient accuracy.

Dr W. H. Eccles: The author has compared the respective approximations obtained by his method and the method of least squares by applying them to a few illustrative examples. It would be useful to attempt some more general comparison. For instance the results obtained by the two methods can be seen by inspection of the coefficients of the normal equations to be closer when the points happen to be scattered parallel to the axis of y than when parallel to the axis of x, which suggests that accidental choice of axes affects the merit of the simpler approximation. I hope the author will extend his study in this way so as to indicate to users those distributions of points that favour his method and those that do not.

Dr N. R. Campbell (communicated): Mr Awbery has not told us why he prefers his method to that of zero sum, to which it is so closely similar. (The only difference is that, while the zero sum method divides the observations into two groups equal in number, Mr Awbery's method divides them according as x is greater or less than $\bar{x}$.) For, compared with the method of zero sum, his method has some obvious disadvantages. It is much more laborious, since the zero sum method requires only his first operation and part of his fourth. Also it is not as directly applicable to a linear equation with more than two variables.

The zero sum method was (according to Whittaker and Robinson) invented by Tobias Mayer in 1748 but seems to have been forgotten for over 100 years. I described it, and (in ignorance of Mayer's work) gave it a name, in *Phil Mag.* **39**, 177 (1920), and have discussed it further in *Phil. Mag.* **47**, 816 (1924) and in my books *Physics* (Camb. Univ. Press, 1920), chap. 17, and *Measurement and Calculation* (Longmans, 1928), chap. 10. It has recently been used by several physicists other than myself in published work.

The results attained by the two methods, and by least squares, are as follows:

Example 1.	Least squares	$R = 46{\cdot}3459\,(1 + 0{\cdot}00391879\,t)$.
	Awbery	$46{\cdot}3463\,(1 + 0{\cdot}00391835\,t)$.
	Zero sum	$46{\cdot}3464\,(1 + 0{\cdot}00391818\,t)$.
Example 2.	Least squares	$y = 96{\cdot}7 - 0{\cdot}444\,x$.
	Awbery	$96{\cdot}8 - 0{\cdot}514\,x$.
	Zero sum	$96{\cdot}6 - 0{\cdot}411\,x$.

In view of the fact that the residuals have an obvious trend in Example 1, so that the assumed formula is only an approximation, there is nothing to suggest that any one of the methods is here more suitable than any other; all are equally consistent and equally inconsistent with the data.

Finally I should like to protest against Mr Awbery's tacit assumption that the accuracy of a method of adjusting observations is to be judged by its agreement with the least square method, which has so often been shown to be entirely devoid of any theoretical foundation.

Author's reply: In reply to Dr Thomas, I think Dr Ferguson has perfectly expressed my view. I agree also with Mr Sears, and indeed might go further. The method of least squares can be misleading if the distribution of errors is not "normal," even though the number of observations is large. However, most experimental work does in fact lead to a distribution which is approximately normal (a fact which can only be found out by experiment) so that the "least square" answer is in most cases the standard by which I would prefer to judge other methods.

Regarding the points raised by the President, it is true that the axes chosen affect the line obtained; in this respect the method resembles that of least squares. Thus if the line representing two points A and B is determined by the method of least squares, it can be either the line AB, or its perpendicular bisector, according to the axes selected. It is clear that the closeness of the results obtained by the present method, and by the method of least squares, cannot be accurately compared in all cases; otherwise the least-square line could be deduced from the present one and the method would be a variant of the method of conducting the arithmetical operations of least squares. It may, however, be stated that the present method gives greater weight to the points more remote from the centre of gravity of the points. It will agree closely with the method of least squares either if the points are closely condensed near the centre, or if they are in any case very accurately on a line. The difference will, I think, increase as these conditions are departed from.

In reply to Dr Campbell: I have not dealt with the general comparison of methods. Had I done so, I should have mentioned the method of zero sum, as well as a very convenient method for those who do not insist on entirely arithmetical modes of procedure. I prefer the method I have given, because the assignment of some definite system of "weighting," of which I can see the rationale, appeals to me.

ARC SPECTRA IN THE REGION λ 1600–2100

By E. W. H. SELWYN, B.Sc.

Communicated by Prof. A. Fowler, F.R.S., March 7, 1929.
Read and discussed May 10, 1929

ABSTRACT. A simple method is described of photographing ordinary arc spectra down to about λ 1600, and measures are given of the spectra thus observed between λ 1600 and λ 2100 of thirteen elements. The measures are believed to be of an accuracy sufficient to make them useful as provisional standards of wave-length. A number of lines hitherto unobserved is recorded and additions have been made to the analysis of the spectra of Mg I, Be I and B I.

During an investigation by Prof. Fowler of the arc spectrum of silicon (Si I) it became necessary to obtain a spectrum of silicon at least approximating to the arc stage and as complete as was reasonably possible in the region λ 1600–2100. The opacity of air effectively limits the spectra of arcs in air to about λ 1850, but advantage was taken of the transparency of nitrogen in devising, originally for the above purpose, the following method, by which regular arc spectra down to about λ 1600 or a little further may be photographed. It was suggested to the author by Prof. Fowler that it would serve a useful purpose to photograph and measure the spectra of the arcs of a number of elements by this method, and these form the subject of the present communication.

A vacuum grating spectrograph (an early model by Messrs Hilger) was used with a fluorite window sealed over the right-hand slit, so that the spectrograph could be exhausted without disturbing any external arrangements. This arrangement, with normal adjustment, gives the region λ 1250–2100 (first order) on a 6 cm. plate. A heavy brass ring with a hole drilled radially into the central aperture and cooled by water circulating around its outer circumference was then fixed to the spectrograph centrally over the window. The arc was arranged to lie as close as possible to the ring and a stream of nitrogen from a cylinder was led into and through the central aperture of the ring as if to blow out the arc. It was found that only a moderate stream of nitrogen was necessary to keep the space between the window and arc relatively transparent, and to blow the dust from the arc away from the window. The nitrogen stream was not rapid enough to cause any inconvenience by blowing out the arc. The spectrograph was exhausted to a pressure of the order of 0·01 mm. of mercury.

It is thought that this method should prove useful for obtaining spectra, due mainly to unionised atoms; for the production of spectra in this region (or at least below λ 1800) of arc type has hitherto only been accomplished by very special methods (as in the case of carbon for instance) or by introducing self-inductance into

the circuit of a spark in hydrogen, which much reduces the intensity of the light. The results here given for boron (B I) indicate the possibilities of usefulness of the method. The method is easy to use but suffers from the disadvantage that the photographs are not generally successful below about λ 1600; but the fact that lines have occasionally been recorded on the plates down to λ 1335 suggests that improvements may be possible which would enable records of arc spectra to be made down to the limit of transparency of fluorite, or possibly, if nitrogen be sufficiently transparent at atmospheric pressure, below that.

Even the short range usefully covered is of interest, however, for two reasons. In the first place, impurity lines due to the elements in the early columns of the periodic table commonly occur in this region; and secondly, the scale on which many previous measures have been based is now under suspicion, and it is a matter of some interest to try and obtain wave-lengths on the true International Ångström scale for this region.

The arcs used were, in general, of the type commonly used in spectroscopy, namely two carbon poles, a small hollow in the lower one carrying a salt of the element whose spectrum was required. Acheson graphite rods were used instead of the ordinary carbons in order to avoid pole impurities. Thus carbon lines were always available as standards for wave-length measurements.

The carbon line, the value for which had previously been accepted as 1931·027 Å.*, was considered, in view of previous experience, of doubtful value as a standard, but it was necessary to use this line, at least in the first reduction of the plates. Fortunately Prof. Fowler had been able to secure measures in the second order of a 10-ft. grating of two silicon lines λλ 2058, 2124, by means of which, adopted as standards, the wave-length of this line was measured as 1930·95 Å., and this value was used in the reduction of the plates. Thus the standard lines for reducing the plates were as shown in Table 1.

Table 1. Standard lines

C	C	N	C	C
λ 1560·27†	λ 1656·27†	λ 1742·74‡	λ 1751·90§	λ 1930·95
1560·66	1657·01	1745·26		
1561·38	1657·37			
	1657·92			
	1658·13			

The uses of these standards only involves the possibility of systematic error in the extrapolation to λ 2100, amounting to about 0·05 Å. at that wave-length, owing to error in measurement of λ 1930·95. Also, when the plates were reduced in this way, the departure from normality ("errors curve") made necessary a correction

* Bowen and Ingram, *Phys. Rev.* **28**, 444 (1926). There is a strong C I line here, but Bowen considers that this is also the position of a C III line. It seems most probable, however, that it is an arc line showing a shift under vacuum spark conditions.

† Bowen, *Phys. Rev.* **29**, 238 (1927).

‡ Bowen and Ingram, *Phys. Rev.* **28**, 446 (1926). This pair of nitrogen lines invariably appeared on the plates.

§ Fowler and Selwyn, *Proc. Roy. Soc.* A, **118**, 42 (1928).

reaching about 0·10 Å. at λ 2100. The proper reduction of the plates therefore required reliable values for the wave-lengths of at least some lines in the neighbourhood of λ 2100.

The method of reduction of the plates was as follows. All the lines were measured and calculated, assuming $\lambda\lambda$ 1656·27, 1930·95 as standards and linear dispersion throughout the plates, the dispersion, of course, being calculated separately from the measures for each plate. The departure from normality could then be estimated over the region λ 1600–1900 with fair accuracy, but up to λ 2100 the errors curve was still rather indeterminate, although its general form was sometimes suggested by the errors below λ 1900. In the case of copper such an extrapolation of the errors curve suggested wave-lengths for the lines between λ 1900 and λ 2100 differing systematically from those given by Eder* for the copper spark. Such a discrepancy had previously been suggested to exist by Shenstone†, who by his analysis of the spectrum Cu II had been able to calculate the wave-lengths of copper lines in this region. As the probable error in these calculated lines was only 0·02 Å., use was made of them to complete the errors curve. The result of this work was to confirm the value adopted for the carbon line (λ 1930·95) and to bring the measurements of the copper lines into conformity with those of Exner and Haschek. Silicon impurity lines provided the means of drawing the errors curve throughout its length for several plates. Except in the cases of strontium and barium, it then became possible by utilising impurity lines to draw errors curves for all plates. The similarity in shape and size of these curves strongly suggests that all measures are on the same scale. There is also, where independent comparison is possible, a general agreement between measures of impurity lines occurring on several plates. The lines are, in effect, measured on Shenstone's calculated copper lines and on Prof. Fowler's silicon lines. Also they show, where comparison is possible, satisfactory agreement with Exner and Haschek's measures. There are grounds therefore for believing that the scale is absolute, and there should be little systematic error in the wave-lengths here given.

Table 2 will reduce Eder's measures to the same scale as is adopted here. It was obtained from a comparison of the different measures on copper, silver, gold and aluminium.

Table 2. Corrections to Eder's measures of Cu lines

λ		λ		λ	
1850	+ 0·08 Å.	1950	+ 0·12 Å.	2050	+ 0·08 Å.
1875	+ 0·08	1975	+ 0·14	2075	+ 0·06
1900	+ 0·08	2000	+ 0·14	2100	+ 0·05
1925	+ 0·10	2025	+ 0·11	2125	+ 0·03

The grating used was of one metre radius and had been ruled at the National Physical Laboratory. Although the dispersion is but 17·5 Å. mm. (in the first order) the definition possible is such that the measurement error on a good line is not likely to be more than 0·03 Å. The resolving power in practice depends only on the slit width and fineness of grain of the plates. It should be possible, theoretically, to

* *Wien. Ber.* **123**, II *a*, 616 (1914). † *Phys. Rev.* **29**, 380 (1927).

separate two lines 0·06 Å. apart at λ 1850—a separation equivalent to one of 0·0035 mm. on the plate and requiring a slit of less than that width. Actually it is a matter of great difficulty to resolve lines 0·20 Å. apart, but pairs of double that separation are usually resolved.

The exposure and development of Schumann plates requires some care if satisfactory results are to be obtained. Without a fair amount of contrast the plates are somewhat difficult to measure, but prolonging development with the object of increasing the contrast often fogs the background and seems to increase the "grain" of the plate. In the ordinary way, the plates were developed with a normal two-solution pyro developer. Occasionally the proportion of accelerator was reduced, but restrainer was not added unless the plate was fogging unduly. In fact it is generally no disadvantage to have some fog present as it ensures that faint lines are developed out. Some of the photographs were taken on fine-grain plates specially prepared by Messrs Hilger; there seems to be a definite advantage in using these plates when the utmost possible resolution is required.

A comparison of the different measures made shows that the accuracy of the wave-lengths here given should be better than 0·05 Å., and it is expected that few lines will be in error by more than this amount.

The following tables give the measures which have been made of the lines of the various spectra. Wave-lengths are given in air down to λ 2000 and in vacuum below that wave-length. It has been more usual to change over at λ 1850, but the publication of Kayser's "Schwingungzahlen," which cease at λ 2000, renders the place adopted here more convenient as regards conversion into wave-numbers. A dotted line indicates where the change over from λ_{air} to $\lambda_{\text{vac.}}$ takes place. In view of the frequent usefulness of such information, details have been given of the series classifications of the lines where this is known.

In the case of beryllium and magnesium new lines have been observed which extend series already known. The plates of boron show doublets extending the series of which the first members were given by Bowen. The intensity and agreement of these lines with the expected positions seems a satisfactory indication of the correctness of their allocation, but, in view of the probable occurrence of doublet and quartet terms other than the regular terms of the series already classified, it would seem desirable to repeat the observations with greater dispersion and resolving power. There are in the measures, it is thought, indications of combinations involving these extra terms.

The sources of the classifications are given at the foot of each table. The notation in all cases except Au I has been brought into line with that which is now generally adopted. In this system the term is written as a capital letter denoting the type (S, P, D, ...), with the multiplicity number as a superscript on the left, and the inner quantum number as a subscript on the right. The prefix denotes the orbit (s, p, d, ...) occupied by the "series" electron (or electrons in some cases), the preceding numeral being the principal quantum number of the orbit. When this prefix is undashed it denotes that one electron from the outermost normal orbit is in an excited state, or that the atom is unexcited. If the prefix symbol is dashed then two

or more electrons are in excited states, or an electron has been raised from an orbit deeper than that which is outermost in the unexcited state of the atom; the symbol then adopted is that of the higher of the orbits occupied by the electrons. A table of some of the more important configurations and corresponding terms of B I is given for illustration.

Table 3. Predicted terms of B I

K	L		M			N				O			Electrons outside rare gas shell	Adopted prefix	Terms
1_1	2_1	2_2	3_1	3_2	3_3	4_1	4_2	4_3	4_4	5_1	5_2	5_3			
2	2	1											$2s^2\ 2p$	$2p$	^2P
2	2		1										$2s^2\ 3s$	$3s$	^2S
2	2			1									$2s^2\ 3p$	$3p$	^2P
2	2				1								$2s^2\ 3d$	$3d$	^2D
2	2					1							$2s^2\ 4s$	$4s$	^2S
2	2						1						$2s^2\ 4p$	$4p$	^2P
2	2							1					$2s^2\ 4d$	$4d$	^2D
2	2								1				$2s^2\ 4f$	$4f$	^2F
2	2									1			$2s^2\ 5s$	$5s$	^2S
2	2										1		$2s^2\ 5p$	$5p$	^2P
2	2											1	$2s^2\ 5d$	$5d$	^2D
2	1	2											$2s\ 2p^2$	$2p'$	$^2S\ ^2P\ ^2D\quad ^4P$
2	1	1	1										$2s\ 2p\ 3s$	$3s'$	$^2P\quad ^4P$
2	1	1		1									$2s\ 2p\ 3p$	$3p'$	$^2S\ ^2P\ ^2D\quad ^4S\ ^4P\ ^4D$

Combinations between terms may be regarded as regulated by the prefix symbols according to the familiar rules, i.e. a term with prefix *s* may combine with a term with prefix *p*, a *p* with *s* and *d*, a *d* with *p* and *f*, etc. Strong combinations are also to be expected between terms with dashed prefixes and those with similar undashed prefixes—*d* with *d'*, *p* with *p'*, etc. The structure of the combinations between terms is regulated by the inner quantum selection rules applied to the *j*-numbers of the terms themselves; namely, $\Delta j = \pm 1$ or 0, with the exception of $0 \rightarrow 0$.

The roman numerals following the wave-numbers of the lines in the tables indicate, in the ordinary way, the state of ionisation of the atom to which they are due: I denotes that the atom is neutral; II, that it is once ionised; and III, that it is twice ionised.

The classification of the lines in Table 4 is due to Shenstone*. The notation has been brought into correspondence with that adopted by Shenstone for Ag II and by McLennan and McLay for Au II, except that the "dashing" or "priming" of certain terms has been omitted as unnecessary; the corresponding prefix for those terms prefixed "b" by Shenstone will be 5*s* (to correspond with the d^9, *s* configuration of the atom).

The classification of Ag II lines is by Shenstone†. Similar results were arrived at by McLennan and McLay‡. The arc doublet is from Fowler§.

* *Loc. cit.*

† *Phys. Rev.* **31**, 317 (1928).

‡ *Trans. Roy. Soc. Can.* **22**, III, 1 (1928).

§ *Report on Series in Line Spectra*, p. 112.

Table 4. Copper lines

λ (Int.)	ν	Classification
2126·05 (10)	47020·8	
2122·95 (9)	47089·4	II $4s\ ^1D_2$–$4p\ ^1D_2$
2117·33 (1)	47214·4	
2112·06 (8)	47332·1	II $4s\ ^1D_2$–$4p\ ^1P_1$
2104·80 (7)	47495·3	II $4s\ ^3D_1$–$4p\ ^3D_2$
2098·41 (0)	47639·9	
2087·98 (0)	47877·8	
2085·33 (1)	47938·6	II $4s\ ^3D_3$–$4p\ ^3F_2$
2079·43 (4*n*)	48074·6	
2054·98 (7)	48646·7	II $4s\ ^3D_2$–$4p\ ^3D_2$
2043·80 (6)	48912·7	II $4s\ ^3D_3$–$4p\ ^3D_3$
2037·13 (6)	49072·8	II $4s\ ^3D_2$–$4p\ ^1F_3$
2035·86 (6)	49103·4	II $4s\ ^3D_1$–$4p\ ^3D_1$
2031·07 (0)	49219·1	
2025·49 (4)	49354·7	II $4s\ ^3D_1$–$4p\ ^1D_2$
2024·40 (5*n*)	49381·5	
2016·89 (1)	49565·3	
2015·61 (1)	49597·8	II $4s\ ^3D_1$–$4p\ ^1P_1$
2012·96 (0)	49662·1	
2000·35 (6)	49991·3	II $4s\ ^3D_3$–$4p\ ^1F_3$
1989·85 (3)	50255·0	II $4s\ ^3D_2$–$4p\ ^3D_1$
1979·97 (4)	50505·8	II $4s\ ^3D_2$–$4p\ ^1D_2$
1970·52 (1)	50748·0	II $4s\ ^3D_2$–$4p\ ^1P_1$
1944·63 (2)	51423·7	II $4s\ ^3D_3$–$4p\ ^1D_2$
1825·42 (2*n*)	54781·0	
1817·35 (0)	55025·2	
1774·91 (2*n*)	56340·9	
1741·63 (1*n*)	57417·5	
1725·51 (1*n*)	57953·9	
1713·46 (1*n*)	58361·4	
1703·84 (0)	58691·0	
1691·10 (0)	59133·1	
1688·80 (00)	59213·6	
1688·03 (00)	59241·0	
1686·78 (0)	59286·3	
1685·52 (0)	59328·9	
1684·80 (0)	59354·2	

Table 5. Silver lines

λ (Int.)	ν	Classification
2125·47 (0*n*)	47033·6	II $5p\ ^3P_2$–$5d\ ^3D_3$
2120·48 (8*n*)	47144·2	II $5s\ ^3D_1$–$5p\ ^1D_2$
2114·77 (0)	47271·5	
2113·87 (10*n*)	47291·6	II $5s\ ^3D_3$–$5p\ ^3D_3$
2075·65 (0*n*)	48162·1	II $5p\ ^3P_1$–$5d\ ^1D_2$
2069·86 (4)	48297·1	I $5s\ ^2S_1$–$6p\ ^2P_1$
2061·19 (4)	48500·2	I $5s\ ^2S_1$–$6p\ ^2P_2$
2033·93 (2)	49150·0	II $5s\ ^3D_2$–$5p\ ^1P_1$
2015·96 (2)	49588·2	II $5s\ ^3D_2$–$5p\ ^3D_1$
2000·72 (3)	49965·2	II $5s\ ^3D_3$–$5p\ ^1F_3$
1994·35 (3)	50141·7	II $5s\ ^3D_2$–$5p\ ^1D_1$
1933·51 (1)	51719·4	II $5s\ ^3D_3$–$5p\ ^1D_2$
1709·33 (0)	58502·5	

Table 6. Gold lines

λ (Int.)	ν	Classification
2082·07 (15*n*)	48013·7	II $6s\ ^3D_3$–$6p\ ^3P_2$
2073·52 (0)	48211·8	
2065·92 (2)	48390·8	II $6s\ ^3D_1$–$6p\ ^3F_2$
2044·60 (4)	48893·5	
2025·49 (7)	49354·7	
2021·38 (7)	49455·3	I 1 $^2D_3$–b_{23}
2014·80 (0)	49616·7	
2012·05 (12*n*)	49684·5	
2000·83 (12*n*)	49963·0	II $6s\ ^3D_3$–$6p\ ^3F_3$
1991·94 (10)	50202·3	
1984·54 (1)	50389·5	
1978·14 (15)	50552·5	
1965·34 (0*n*)	50881·8	
1951·97 (7)	51230·3	I 1 $^2S_1$–2 $^2P_2$
1944·72 (3)	51421·3	
1939·24 (5)	51566·6	I 1 $^2D_3$–3 $^2P_2$
1925·26 (3)	51941·0	II $6p\ ^3F_3$–$6d\ ^3G_4$
1921·67 (6)	52038·1	II $6s\ ^1D_2$–$6p\ ^1P_1$
1919·63 (8)	52093·4	I 1 $^2D_3$–c_{23}
1904·55 (0)	52505·8	II $6p\ ^3F_3$–$6d\ ^3D_3$
1879·87 (6)	53195·2	I 1 $^2S_1$–2 $^2P_1$
1872·42 (6)	53406·0	
1857·22 (5)	53843·9	
1855·54 (4)	53892·7	II $6p\ ^3P_2$–$6p\ ^1P_1$
1854·71 (0)	53916·8	
1851·53 (0)	54009·4	II $6p\ ^3P_2$–$6d\ ^3P_2$
1834·28 (5)	54517·3	
1833·14 (5)	54551·2	I 1 $^2D_3$–d_{23}
1823·27 (4)	54846·5	II $6s\ ^3D_1$–$6p\ ^3P_0$
1800·62 (7)	55536·4	II $6s\ ^3D_2$–$6p\ ^3D_2$
1793·31 (7)	55762·8	II $6s\ ^3D_2$–$6p\ ^3P_1$
1783·18 (7)	56079·6	II $6s\ ^1D_2$–$6p\ ^1F_3$
1775·20 (0)	56331·7	
1756·13 (2)	56943·4	II $6s\ ^1D_2$–$6p\ ^1D_2$
1749·78 (4)	57150·0	II $6s\ ^3D_2$–$6p\ ^3D_3$
1740·50 (4)	57454·8	II $6s\ ^3D_3$–$6p\ ^3F_4$
1725·91 (2)	57940·4	II $6s\ ^3D_1$–$6p\ ^3D_1$
1720·03 (1)	58138·5	II $6s\ ^3D_3$–$6p\ ^3D_2$
1718·27 (2)	58198·1	
1716·96 (0)	58242·5	
1711·47 (0*n*)	58428·3	
1673·60 (3)	59751·4	II $6s\ ^3D_3$–$6p\ ^3D_1$
1670·82 (1)	59850·5	
1665·73 (1)	60033·7	I 1 $^2S_1$–3 $^2P_1$
1646·66 (0)	60729·0	I 1 $^2S_1$–3 $^2P_2$

The classification in Table 6 is by M'Lennan and M'Lay*. It has been considered advisable not to change the notation adopted by the authors of the analysis of Au I owing to the inclusion of a number of terms whose relation to the atomic configuration is not clear.

Table 7. Beryllium lines

λ (Int.)	ν	Classification	λ (Int.)	ν	Classification
*2125·70 (5)	47028·5	I $2p\ ^3P_{210}-5s\ ^3S_1$	1956·97 (1*n*)	51099·4	I $2p\ ^3P_{210}-8s\ ^3S_1$
*2056·06 (9*n*)	48621·1	I $2p\ ^3P_{210}-5d\ ^3D_{321}$	*1943·84 (4*n*)	51444·6	I $2p\ ^3P_{210}-8d\ ^3D_{321}$
2050·45 (2*n*)	48754·1		1929·47 (1*n*)	51827·7	I $2p\ ^3P_{210}-9d\ ^3D_{321}$
2043·13 (1*n*)	48928·7		1919·62 (1*n*)	52093·3	I $2p\ ^3P_{210}-10d\ ^3D_{321}$
*2032·80 (5)	49177·3	I $2p\ ^3P_{210}-6s\ ^3S_1$	1912·75 (0*n*)	52280·7	I $2p\ ^3P_{210}-11d\ ^3D_{321}$
*1998·11 (7*n*)	50047·3	I $2p\ ^3P_{210}-6d\ ^3D_{321}$	*1776·28 (6)	56297·4	II $2p\ ^2P_{21}-3s\ ^2S_1$
1985·34 (2)	50369·2	I $2p\ ^3P_{210}-7s\ ^3S_1$	1721·43 (1)	58091·2	? Al.
*1964·79 (6*n*)	50896·0	I $2p\ ^3P_{210}-7d\ ^3D_{321}$	1661·48 (2)	60187·3	

Classifications of lines marked by an asterisk in Table 7 are by Bowen and Millikan†. The rest of the lines classified by them in Be I are given in Table 8.

Table 8

λ_{air}	ν	Classification	λ_{air}	ν	Classification
3321·350	30099·60	$2p\ ^3P_2-3s\ ^3S_1$	2494·734	40072·34	$2p\ ^3P_2-3d\ ^3D_{321}$
3321·089	30101·97	$2p\ ^3P_1-3s\ ^3S_1$	2494·589	40074·67	$2p\ ^3P_1-3d\ ^3D_{21}$
3321·016	30102·63	$2p\ ^3P_0-3s\ ^3S_1$	2494·546	40075·36	$2p\ ^3P_0-3d\ ^3D_1$
2650·776	37713·59	$2p\ ^3P_2-2p'\ ^3P_1$	2350·78	42526·0	$2p\ ^3P_{210}-4s\ ^3S_1$
2650·719	37714·54	$2p\ ^3P_1-2p'\ ^3P_0$	2175·04	45961·8	$2p\ ^3P_{210}-4d\ ^3D_{321}$
2650·645	37715·45	$2p\ ^3P_2-2p'\ ^3P_2$			
2650·618	37715·84	$2p\ ^3P_1-2p'\ ^3P_1$			
2650·562	37716·63	$2p\ ^3P_0-2p'\ ^3P_1$			
2650·467	37717·99	$2p\ ^3P_1-2p'\ ^3P_2$			

The following formulae have been computed for the above series:

$$2p\,^3P_1-(m+2)\,s\,^3S_1=53199\cdot8-\frac{R}{[m+1\cdot237248-0\cdot057234\,m]^2}\qquad R=109737\quad m=1,4,6.$$

$$2p\,^3P_1-(m+1)\,d\,^3D=53224\cdot8-\frac{R}{[m+0\cdot836383+0\cdot104715/m]^2}\qquad R=109737\quad m=2,7,9.$$

There is no reason therefore for any alteration of the value 53212·86 adopted by Millikan and Bowen‡ for the term $2p\ ^3P_0$. With the use of this value the terms of the Be I spectrum are given by the above observations as follows:

Table 9. Terms of Be I

$2p\ ^3P_0$	53212·86	$3s\ ^3S_1$	23110·22	$3d\ ^3D_{321}$	13137·50
$2p\ ^3P_1$	53212·18	$4s\ ^3S_1$	10685·1	$4d\ ^3D_{321}$	7249·3
$2p\ ^3P_2$	53209·83	$5s\ ^3S_1$	6182·6	$5d\ ^3D_{321}$	4590·0
		$6s\ ^3S_1$	4033·8	$6d\ ^3D_{321}$	3163·8
		$7s\ ^3S_1$	2841·9	$7d\ ^3D_{321}$	2315·1
		$8s\ ^3S_1$	2111·7	$8d\ ^3D_{321}$	1766·5
$2p'\ ^3P_0$	15497·64			$9d\ ^3D_{321}$	1383·4
$2p'\ ^3P_1$	15496·27			$10d\ ^3D_{321}$	1117·8
$2p'\ ^3P_2$	15494·28			$11d\ ^3D_{321}$	930·4

* Au I, *Proc. Roy. Soc.* A, **112**, 95 (1926); Au II, *Trans. Roy. Soc. Can.* **22**, III, 103 (1928).
† *Phys. Rev.* **28**, 256 (1926). ‡ *Loc. cit.*

The *s*-terms have thus been extended in Table 9 by two and the *d*-terms by three additional members.

Table 10. Magnesium

λ (Int.)	ν	Classification	λ (Int.)	ν	Classification
2025·86 (15*n*)	49346·7	I $3s\ {}^1S_0 - 4p\ {}^1P_1$	1707·10 (3*n*)	58578·9	I $3s\ {}^1S_0 - 7p\ {}^1P_1$
1827·97 (8*n*)	54705·5	I $3s\ {}^1S_0 - 5p\ {}^1P_1$	1683·51 (1*n*)	59399·7	I $3s\ {}^1S_0 - 8p\ {}^1P_1$
1747·81 (5*n*)	57214·5	I $3s\ {}^1S_0 - 6p\ {}^1P_1$	1668·57 (0*n*)	59931·6	I $3s\ {}^1S_0 - 9p\ {}^1P_1$

The first two lines given in Table 10 are classified by Fowler*; the remainder have not previously been recorded. The following formula has been computed from the first, second and fifth lines given above. The series is thus represented (with the exception of the first line given in Fowler's *Report*, which is 90 wave-number units out of its calculated position) within about 1 wave-number unit:

$$3s\,{}^1S_0 - (m+2)\,p\,{}^1P_1 = 61668 - \frac{R}{[m + 0{\cdot}941713 + 0{\cdot}085341/m]^2} \qquad \begin{array}{l} R = 109737 \\ m = 1, 2, 3, \ldots\ldots \end{array}$$

The limit thus calculated is 4 units lower than that adopted in Fowler's *Report*. With the limit in the *Report* the above observations yield the terms and Rydberg denominators shown in Table 11.

Table 11. Terms of Mg I

Old Notation	New Notation	Term Value	Rydberg Denominator
2P	$4p\ {}^1P_1$	12325·4	2·9838
3P	$5p\ {}^1P_1$	6966·6	3·9689
4P	$6p\ {}^1P_1$	4457·6	4·9616
5P	$7p\ {}^1P_1$	3093·2	5·9563
6P	$8p\ {}^1P_1$	2272·4	6·9492
7P	$9p\ {}^1P_1$	1740·5	7·9403

Table 12. Calcium lines

λ (Int.)	ν	Classification	λ (Int.)	ν	Classification
2112·77 (10)	47316·2	II $4p\ {}^2P_2 - 5d\ {}^2D_{32}$	1814·59 (1)	55108·9	II $4p\ {}^2P_2 - 6d\ {}^2D_3$
2103·24 (8)	47530·6	II $4p\ {}^2P_1 - 5d\ {}^2D_2$	1807·39 (1)	55328·4	II $4p\ {}^2P_1 - 6d\ {}^2D_2$
1840·07 (8)	54345·8	II $4d\ {}^2D_3 - 4f\ {}^2F$	1651·99 (1)	60533·1	II $4s\ {}^2S_1 - 5p\ {}^2P_1$
1838·04 (7)	54405·8	II $3d\ {}^2D_2 - 4f\ {}^2F$	1649·87 (1)	60610·8	II $4s\ {}^2S_1 - 5p\ {}^2P_2$

Table 13. Strontium lines

λ (Int.)	ν	Classification	λ (Int.)	ν	Classification
2124·48 (4)	47055·5		1778·51 (7)	56226·8	II $4d\ {}^2D_3 - 6f\ {}^2F$
2052·18 (3)	48713·0	II $5p\ {}^2P_2 - 8s\ {}^2S_1$?	1769·65 (6)	56508·4	II $4d\ {}^2D_2 - 6f\ {}^2F$
1995·24 (5)	50119·3	II $5p\ {}^2P_2 - 7d\ {}^2D_3$	1620·35 (1)	61715·1	II $4d\ {}^2D_3 - 7f\ {}^2F$
1964·67 (3)	50899·1	II $5p\ {}^2P_1 - 7d\ {}^2D_2$	1612·98 (0)	61997·0	II $4d\ {}^2D_2 - 7f\ {}^2F$

* *Report on Series in Line Spectra*, p. 117.

Table 14. Barium lines

λ (Int.)	ν	Classification	λ (Int.)	ν	Classification
2052·97 (2)	48695·0	II $6p\ ^2P_2-9d\ ^2D_3$	1904·16 (1)	52516·6	
2024·18 (4)	49386·9		1786·76 (1)	55967·2	
1999·54 (3)	50011·5		1694·31 (4)	59021·1	
1985·89 (1)	50355·3	II $6p\ ^2P_1-9d\ ^2D_2$	1674·39 (1)	59723·2	
1924·77 (5)	51954·3				

The classifications in Tables 12, 13 and 14 are all from Fowler's *Report**, except $4s\ ^2S-5p\ ^2P$ of Ca II which was recognised by Saunders and Russell†. In the case of strontium and barium there may be a small systematic error in the wave-lengths—a correction was made by fitting an errors curve, typical of the other plates measured, to the errors below λ 1930·95. The wave-length scale should therefore be correct at λ 1600, any systematic error being proportional to the increase in wave-length above 1600 Å.

Table 15. Zinc lines

λ (Int.)	ν	Classification	λ (Int.)	ν	Classification
2115·04 (2)	47265·5		2025·49 (15)	49354·7	II $4s\ ^2S-4p\ ^2P$
2111·85 (0)	47336·8		2022·05 (1)	49438·9	
2104·39 (7)	47504·6	I $4p\ ^3P_2-4p'\ ^3P_1$	1972·46 (2)	50698·1	
2102·19 (2)	47554·3	II $4p\ ^2P-4d\ ^2D$	1921·05 (1)	52054·9	
2099·93 (6)	47605·5	II $4p\ ^2P-4d\ ^2D$	1904·82 (2)	52498·4	
2096·89 (5)	47674·4	I $4p\ ^3P_1-4p'\ ^3P_0$	1898·92 (1)	52661·5	
2087·12 (6*n*)	47897·5	I $\begin{cases}4p\ ^3P_1-4p'\ ^3P_1\\4p\ ^3P_2-4p'\ ^3P_2\,?\end{cases}$	1868·82 (1)	53509·7	
2079·02 (6)	48084·1	I $4p\ ^3P_0-4p'\ ^3P_1$	1822·05 (5)	54883·2	
2070·08 (4*n*)	48291·9	I $4p\ ^3P_1-4p'\ ^3P_2$?	1813·04 (1)	55156·0	
2064·23 (2)	48428·7	II $4p\ ^2P-4d\ ^2D$	1796·68 (2)	55658·2	
2062·01 (10)	48480·9	II $4p\ ^2S_1-4p\ ^2P$	1726·79 (1)	57910·9	
2060·03 (1*n*)	48527·4		1682·11 (5)	59449·2	
2053·27 (2)	48687·2		1671·54 (3)	59825·1	
2050·92 (0*n*)	48742·9				

Classifications of Zn II lines are due to von Salis‡. The fragmentary PP′ group is from Grotrian§. The lines marked ? are suggested as completing the group.

Table 16. Cadmium lines

λ (Int.)	ν	Classification	λ (Int.)	ν	Classification
1820·37 (0)	54933·9		1669·02 (1*n*)	59915·4	
1669·45 (0*n*)	59900·0				

Table 17. Mercury lines

λ (Int.)	ν	Classification	λ (Int.)	ν	Classification
2052·81 (1)	48698·1		1849·68 (20R)	54063·4	I $6s\ ^1S_0-6p\ ^1P_1$
2001·99 (2*n*)	49934·1	I $6p\ ^3P_2-6p'\ ^3P_1$	1832·74 (1*n*)	54563·1	I $6p\ ^3P_1-6p'\ ^3P_1$
1972·94 (3)	50685·8		1775·20 (2*n*)	56331·7	I $6p\ ^3P_0-6p'\ ^3P_1$
1942·28 (5)	51485·9	II $6s\ ^2S-6p\ ^2P$	1649·90 (3)	60609·7	II $6s\ ^2S-6p\ ^2P$
1869·25 (0)	53497·4				

* Ca II, p. 127; Sr II, p. 132; Ba II, p. 137.
† *Astrophys. Journ.* **62**, 5 (1925).
‡ *Ann. der Phys.* **76**, 145 (1925).
§ *Graphische Darstellung der Spectren*, **2**, 151 (1928).

The PP′ group in Table 17 is again from Grotrian*, the singlet λ 1849·68 from Fowler's *Report*†. It is notable that this measure does not agree so well with the position calculated from the triplet-singlet inter-combinations as the earlier measure by Wolff. The lines of Hg II are as classified by Carroll‡.

Table 18. Boron lines

λ (Int.)	ν	Classification	λ (Int.)	ν	Classification
*2089·57 (10*n*)	47841·4	I $2p\,^2P_2-2p'\,^2D_3$	1663·07 (2)	60129·8	I $2p\,^2P_2-5s\,^2S_1$
*2088·84 (9*n*)	47858·1	I $2p\,^2P_1-2p'\,^2D_2$	1662·62 (1)	60146·0	I $2p\,^2P_1-5s\,^2S_1$
*2067·17 (6)	48359·9	{I III $2s\,^2S_1-2p\,^2P_1$	*1624·38 (0)	61562·0	II $2p\,^3P_{210}-2p'\,^3P_{210}$
2066·61 (7*n*)	48373·0	I	*1624·06 (3)	61574·1	
*2065·78 (1)	48392·4	III $2s\,^2S_1-2p\,^2P_2$	*1623·61 (1)	61591·2	
*1842·85 (1)	54263·8	II $2p\,^1P_1-3s\,^1S_0$	1607·78 (0)	62197·6	
*1826·52 (9*n*)	54748·9	I $2p\,^2P_2-3d\,^2D_{32}$	1600·91 (1*n*)	62464·5	I $2p\,^2P_{21}-5d\,^2D_{32}$
*1825·97 (8*n*)	54765·4	I $2p\,^2P_1-3d\,^2D_2$	1600·28 (0*n*)	62489·1	
1818·41 (6)	54993·1	I $2p\,^2P_2-4s\,^2S_1$	1573·44 (2*n*)	63555·0	
1817·90 (5)	55008·5	I $2p\,^2P_1-4s\,^2S_1$	1566·64 (1*n*)	63830·9	I $2p\,^2P_{21}-6d\,^2D_{32}$
1667·42 (3*n*)	59972·9	I $2p\,^2P_2-4d\,^2D_{32}$	1558·93 (0*n*)	64146·6	
1666·99 (2*n*)	59988·4	I $2p\,^2P_1-4d\,^2D_2$			

Boric acid (H_2BO_3) was used on carbon poles as the source of the spectrum detailed in Table 18. Classifications of lines marked by an asterisk are by Bowen and Millikan§ and Bowen‖.

The first three pairs $2p\,^2P_{21}-(m+1)\,d\,^2D_{321}$ give the following series formula:

$$\nu = \begin{matrix}66911\\66927\end{matrix} - \frac{R}{[m+0{\cdot}923463+0{\cdot}160716/m]^2} \qquad \begin{matrix}R = 109737\\ m = 2, 3, 4, \ldots\ldots\end{matrix}$$

and the $2p\,^2P_{21}-(m+2)\,s\,^2S_1$ pairs the following (using λλ 2497·73, 2496·78, νν 40024, 40040, given in the *Report*, p. 155, as $2p\,^2P_{21}-3s\,^2S_1$):

$$\nu = \begin{matrix}66796\\66812\end{matrix} - \frac{R}{[m+1{\cdot}073754-0{\cdot}049175/m]^2} \qquad \begin{matrix}R = 109737\\ m = 1, 2, 3, \ldots\ldots\end{matrix}$$

The adoption of 66840 as the value of $2p\,^2P_1$ gives the terms of the B I spectrum shown in Table 19.

Table 19. Terms of B I

$2p\,^2P_1$	66840	$3s\,^2S_1$	26800	$3d\,^2D_{32}$	12075
$2p\,^2P_2$	66824	$4s\,^2S_1$	11831	$4d\,^2D_{32}$	6851
		$5s\,^2S_1$	6694	$5d\,^2D_{32}$	4359
$2p'\,^2D_{32}$	18983			$6d\,^2D_{32}$	2993

The wave-lengths given in Table 20 were measured by copper lines introduced for the purpose in addition to the carbon lines, and agree very well with those given by Bowen and Ingram¶. The PP′ group of Al I (one line being obscured) was given by Bowen and Millikan**. The classification of Al II lines is by Paschen and Sawyer††, and of Al III by Paschen‡‡. The strong diffuse pair, λλ 1932, 1936, is of some interest

* *Loc. cit.* † p. 149. ‡ *Phil. Trans.* A, **225**, 366 (1926).

§ B II, *Phys. Rev.* **26**, 310 (1925); B III, *Proc. Nat. Acad. Amer.* **10**, 199 (1924).

‖ B I, *Phys. Rev.* **29**, 231 (1927).

¶ *Phys. Rev.* **28**, p. 444 (1926).

** *Phys. Rev.* **26**, 160 (1925).

†† *Ann. der Phys.* **84**, 16 (1927).

‡‡ *Ann. der Phys.* **71**, 537 (1923).

Table 20. Aluminium lines

λ (Int.)	ν	Classification	λ (Int.)	ν	Classification
1990·56 (2)	50237·1	II $3p\ ^1P_1-4d\ ^1D_2$	1766·41 (2)	56612·0	I $3p\ ^2P_2-3p'\ ^2P_2$
1936·45 (12n)	51640·9		1765·81 (1)	56631·2	II $3p\ ^3P_1-3p'\ ^3P_0$
1932·25 (10n)	51753·1		1763·95 (2)	56691·0	II $3p\ ^3P_{21}-3p'\ ^3P_{21}$
1862·75 (1)	53684·1	III $3s\ ^2S_1-3p\ ^2P_1$	1762·97 (1n)	56722·5	I $3p\ ^2P_1-3p'\ ^2P_2$
1862·42 (1)	53693·6	II $3p\ ^3P_2-4s\ ^3S_1$	1762·03 (0)	56752·7	II $3p\ ^3P_0-3p'\ ^3P_1$
1858·10 (1)	53818·4	II $3p\ ^3P_1-4s\ ^3S_1$	1760·13 (1)	56814·0	II $3p\ ^3P_1-3p'\ ^3P_2$
1856·01 (0)	53879·0	II $3p\ ^3P_0-4s\ ^3S_1$	1724·99 (3)	57971·4	II $3p\ ^3P_2-3d\ ^3D$
1854·75 (1)	53915·6	II $4s\ ^3S_1-9p\ ^3P$	1721·28 (2)	58096·3	II $3p\ ^3P_1-3d\ ^3D$
		III $3s\ ^2S_1-3p\ ^2P_2$	1719·46 (1)	58157·8	II $3p\ ^3P_0-3d\ ^3D$
1769·19 (1)	56523·1	I $3p\ ^2P_2-3p'\ ^2P_1$	1670·76 (4n)	59853·0	II $3p\ ^1S_0-3p\ ^1P_1$
1767·77 (2)	56568·5	II $3p\ ^3P_2-3p'\ ^3P_1$			

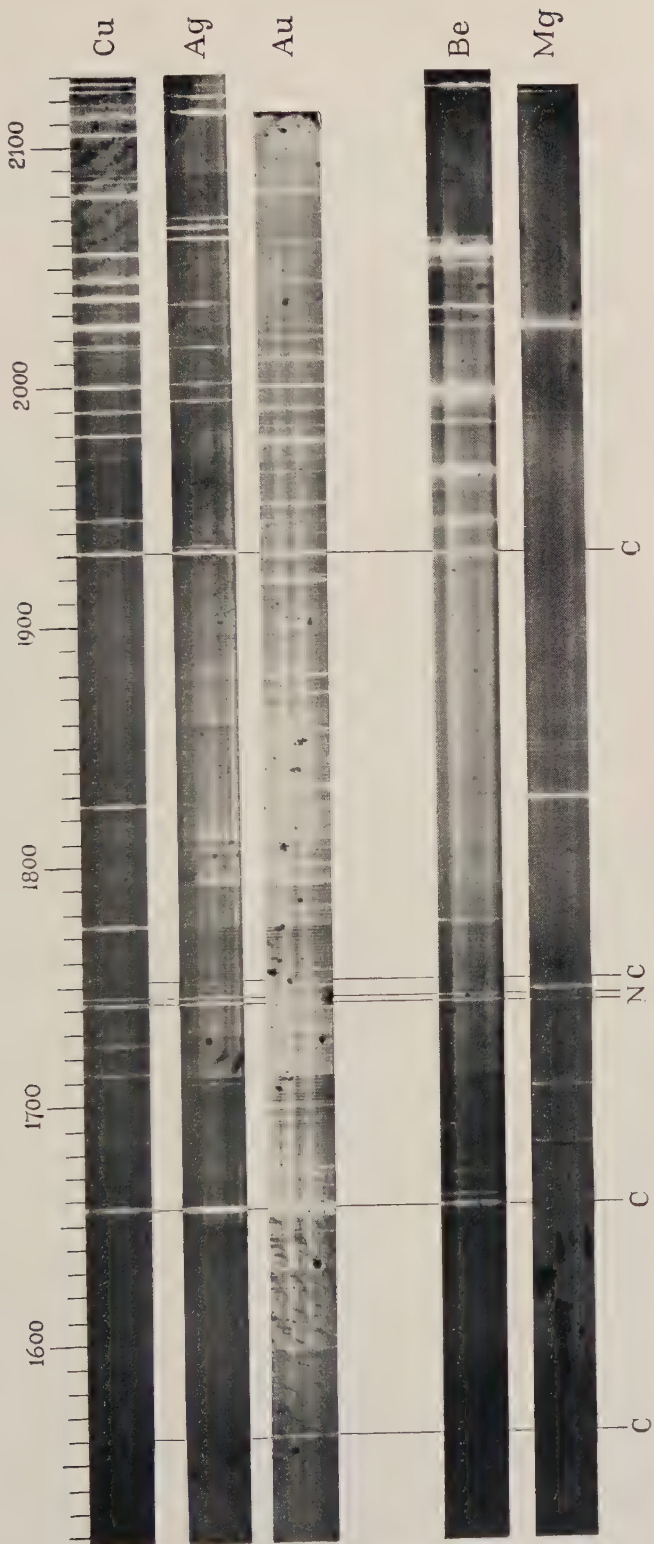
as it has the separation of $\Delta\nu$ 112 of Al I exactly; Paschen and Sawyer allot a number of lines here (λλ 1930·03, 1932·43, 1934·54, 1936·96) to a group $^3P'\ ^3P''$ of Al II. It seems probable that this pair, which is evidently an arc pair, arises from the combination $3p\ ^2P_{21}-3p'\ ^2S_1$, the latter term being associated with the electronic configuration sp^2.

Table 21 is of lines which have appeared as impurities on the plates taken during this investigation. It is thought that this table may be of use in the elimination of impurity lines in spectra in the region concerned. Wave-lengths are given *in vacuo* over the whole range and in air down to λ 1800.

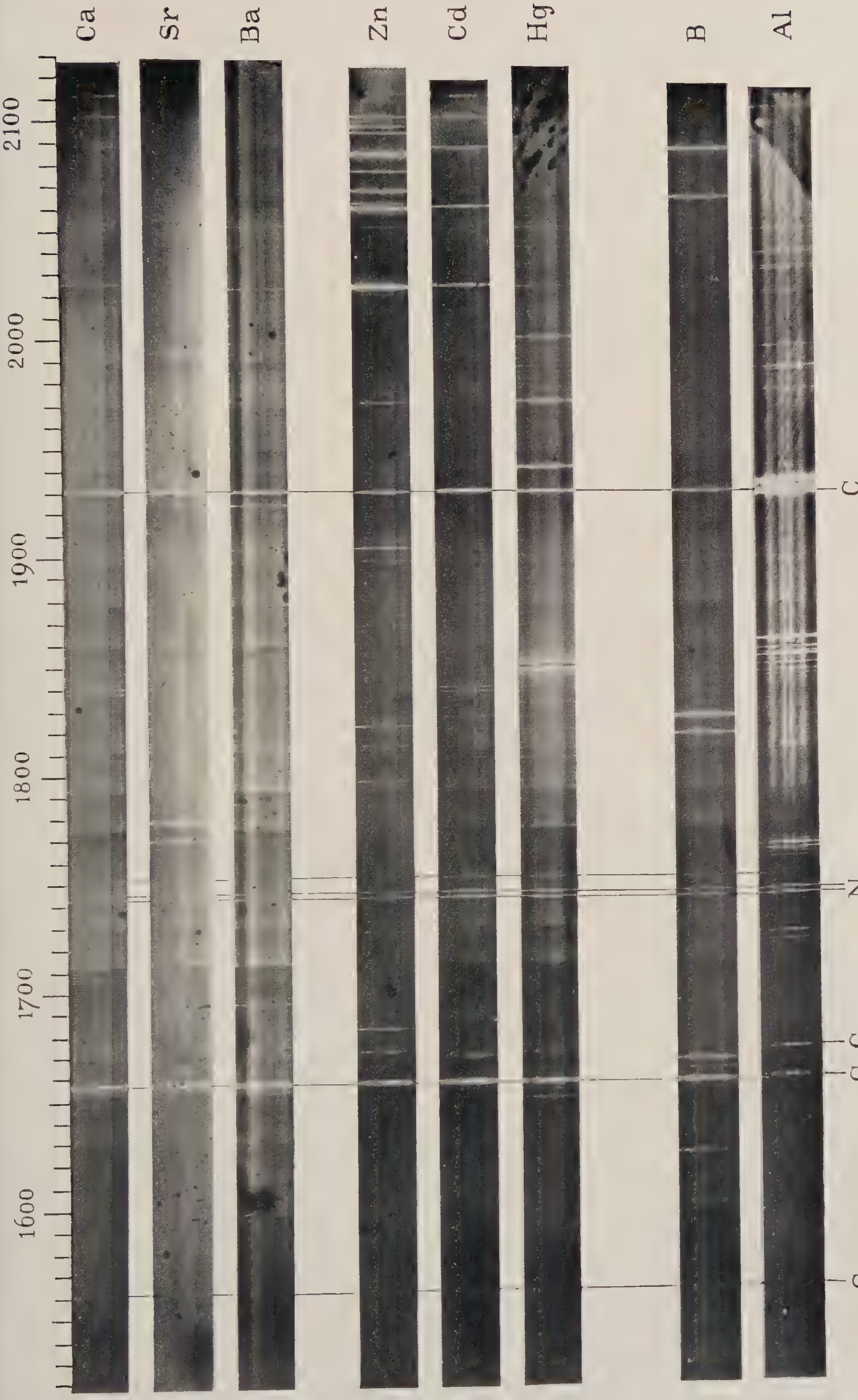
Table 21. List of impurity lines

λ_{air}	$\lambda_{vac.}$	Origin	λ_{air}	$\lambda_{vac.}$	Origin
2124·150 (10R)	2124·824	Si I	1935·81 (12n)	1936·45	Al
2089·57 (10n)	2090·24	B I	1931·61 (10n)	1932·25	,,
2088·84 (9n)	2089·51	,,	1900·73 (8R)	1901·36	Si I
2069·86 (4)	2070·52	Ag I	1851·87 (2)	1852·49	,,
2062·01 (10)	2062·67	Zn II	1850·08 (4)	1850·70	,,
2061·19 (4)	2061·85	Ag I	1847·55 (2)	1848·17	,,
2058·139 (5)	2058·801	Si I	1846·86 (3)	1847·48	,,
2054·98 (7)	2055·64	Cu II	1844·91 (3)	1845·53	,,
2044·60 (4)	2045·26	Au	1840·86 (3)	1841·48	,,
2043·80 (6)	2044·46	Cu II	1839·45 (8)	1840·07	Ca II
1037·13 (6)	2037·79	,,	1837·42 (7)	1838·04	,,
2035·86 (6)	2036·52	,,	1827·35 (8n)	1827·97	Mg I
2025·86 (15n)	2026·52	Mg I	1825·90 (9n)	1826·52	B I
2025·49 (15)	2026·15	Zn II	1825·35 (8n)	1825·97	,,
2024·40 (5n)	2025·05	Cu		1769·19 (1)	Al I
2015·96 (2)	2016·61	Ag II		1766·41 (2)	,,
1999·70 (6)	2000·35	Cu II		1765·81 (1)	Al II
1993·70 (3)	1994·35	Ag II		1763·95 (2)	,,
1988·36 (6)	1989·01	Si I		1762·97 (1n)	Al I
1985·74 (4)	1986·39	,,		1724·99 (3)	Al II
1982·60 (3)	1983·25	,,		1721·28 (2)	,,
1978·58 (3)	1979·23	,,		1719·46 (1)	,,
1976·97 (3)	1977·62	,,		1670·76 (4n)	,,
1943·99 (2)	1944·63	Cu II			

I am indebted to Prof. Fowler for permission to use his measures of silicon lines before publication, and I wish to express my appreciation of the help and encouragement I have received from him.



Arc spectra λ 1600–2100



Arc spectra λ 1600–2100

Note added May 10, 1929. The investigation of Si I mentioned at the commence-ment will be found described in *Proc. Roy. Soc.* A, **123**, 422 (1929).

DESCRIPTION OF PLATES

The plates contain reproductions of photographs taken in this investigation; the following details may be of interest.

		Source	Exposure (min.)
Cu	Copper	Metal on carbon poles	10
Ag	Silver	Metal on carbon poles	10
Au	Gold	Metal on carbon poles	15
Be	Beryllium	Metal on carbon poles	10
Mg	Magnesium	Metallic magnesium poles	7
Ca	Calcium	Chloride on carbon poles	10
Sr	Strontium	Chloride on carbon poles	15
Ba	Barium	Hydroxide on carbon poles	20
Zn	Zinc	Metallic zinc and carbon poles	15
Cd	Cadmium	Chloride on carbon poles	20
Hg	Mercury	Mercuric oxide on carbon poles	15
B	Boron	Boric acid on carbon poles	7
Al	Aluminium	Metallic aluminium, copper and carbon poles	10

DISCUSSION

Dr K. R. Rao: I have used the method described by Mr Selwyn in the course of my investigations on the arc spectra of germanium and arsenic and I wish to record that I have found it peculiarly suitable for the study of arc spectra in the region λλ 2000–1560 where it is difficult by the usual methods (e.g. the spark between metallic poles) to differentiate the arc lines from those of the spark.

Mr L. J. Freeman: In connection with the appearance of lines down to λ 1335, could Mr Selwyn tell us whether the transparency of nitrogen was apparently continuous down to that wave-length?

Author's reply: Owing to the formation of oxides of nitrogen by the arc there is some discontinuous absorption over the range for which results are given, and it is possible that this discontinuous absorption continues to regions of shorter wave-length. But nitrogen itself appears to be continuously transparent, or rather, any absorption bands to which it may give rise are comparatively weak and broad. Judging from my own experience, if oxygen could be entirely eliminated from the neighbourhood of the arc, photographs would be secured down to the limit of transparency of fluorite.

HEAVISIDE'S FRACTIONAL DIFFERENTIATOR

By W. E. SUMPNER, D.Sc., M.Inst.E.E.

Received November 21, 1928 *and in revised form March* 5, 1929. *Read May* 10, 1929

ABSTRACT. Operational methods involve the assumption that the operator may be treated whenever convenient either as an algebraic quantity or as the differentiator of the calculus. They are justified if theorems can be found to cover the processes actually used. Such theorems were simple and often obvious in the early methods of Boole and his immediate successors. In later and more elaborate methods the necessary theorems are usually assumed without proof, so that these methods are only justified by results.

Heaviside fearlessly extended such methods in dealing with the theory of electromagnetic waves. In particular he used fractional powers of p, though no one has found a meaning for a fractional power of the differentiator d/dx.

Several important papers have been published, notably those by Bromwich, Jeffreys, and Carson, dealing with Heaviside's work; but it is not unfair to say that they set forth independent parallel methods justifying the results of Heaviside, rather than any explanation of his methods.

In the present paper the operator p is defined in accordance with Heaviside's rule, but no assumption is made that it is identical with the differentiator of the calculus.

The mathematical functions used in physics consist of series of terms involving powers of the variable t

$$\Sigma\, a_n t^n \quad \text{or} \quad \Sigma\, c_n \frac{t^n}{n!},$$

where a_n and c_n are constants and $t_n\, n!$ is a power term whose index is n. The operator p^a is defined simply as an order to reduce by a the index of each power term to which the operator is applied, thus

$$p^a \Sigma\, c_n . \frac{t^n}{n!} = \Sigma\, c_n \frac{t^{n-a}}{(n-a)!},$$

including

$$p^a . 1 = p^a \frac{t^0}{0!} = \frac{t^{-a}}{(-a)!}.$$

A differential equation $d^2V/dx^2 = m^2 . dV/dt$ can be regarded as an index relation

$$\frac{d^2V}{dx^2} = m^2 p^1 V \quad \text{or} \quad \left(\frac{d}{dx} - mp^{\frac{1}{2}}\right)\left(\frac{d}{dx} + mp^{\frac{1}{2}}\right) V = 0,$$

by means of which a connection can be found between the co-efficients of a power series in t assumed for V. If we can find two power series P and Q such that

$$\frac{dP}{dx} = mp^{\frac{1}{2}} P \quad \text{and} \quad \frac{dQ}{dx} = -\, mp^{\frac{1}{2}} Q,$$

the solution for V will be

$$V = AP + BQ,$$

where A and B are constants.

The power series denoted by

$$P = e^{xmp^{\frac{1}{2}}}, \quad Q = e^{-xmp^{\frac{1}{2}}},$$

or

$$P = 1 + xm \frac{t^{-\frac{1}{2}}}{(-\frac{1}{2})} + \frac{x^2 m^2}{2!} \frac{t^{-1}}{(-1)!} + \text{etc.},$$

with a similar expression for Q, will fulfil all necessary conditions.

In finding this solution p^a has been used merely as an index changer without assuming any connection with differentiation.

By proceeding on these lines, and by using Leibnitz's theorem to suggest an operator for fractional powers of p, it proves possible to establish theorems justifying many of Heaviside's methods.

§ 1. HEAVISIDE'S EXPERIMENTAL METHODS

HEAVISIDE'S mathematical writings are of special interest to the experimental worker in physics. The problems dealt with are of importance in electrical science, and the methods and results have been of great service in electrical technology. Moreover Heaviside's whole outlook upon mathematics was that of an experimenter.

He was "emphatic on the matter of the use of experiment in mathematics even without proper understanding." "It is just the same in mathematical research into unknown regions as in experimental physical research." Theory "can only become logical when the subject is very well known indeed." "Facts and experiments come first."

Fresh work in mathematics, as in other sciences, must be experimental in its early stages. Heaviside held the view that in mathematical research there is a strong tendency to standardise and to become conventional too soon, so that progress is less rapid than might otherwise be the case. No one was ever less willing to follow fashion than Heaviside. He was more interested in devising new methods to get at results quickly than in finding firm proofs of such new methods. He was convinced about results as soon as he could verify them by severe experimental tests, and passed on without waiting to find formal proofs. He was a wanderer in the wilds and loved country far beyond railhead.

A large part of his work on electromagnetic theory relates to the behaviour of waves. His methods involved operators the action of which he interpreted with the aid of his fractional differentiator. This was used merely as a tool in connection with his method, and is to be distinguished from the method itself. He made great use of the tool, but did not profess to explain it, and never regarded the theory of it as one of pressing importance. Criticism of his work has in general referred not so much to his methods as to the use of the fractional operator in conjunction with them, and has been based on the assumption that this operator is identical with the differentiator of the calculus. In what follows I shall try to show that, in order to justify the use of the operator, it is not at all necessary to assume this identity in meaning*.

§ 2. THE INDEX OPERATOR, ITS DEFINITION AND JUSTIFICATION

Heaviside's fractional differentiator, which is perhaps better described as an index operator, can be defined by the equations

$$p^{\alpha}t^{\beta} = \beta!\, p^{\alpha}(t^{\beta}/\beta!) = \beta!\,[t^{\beta-\alpha}/(\beta-\alpha)!] \qquad \text{......(1),}$$

* In papers by Bromwich, *Proc. Lond. Math. Soc.* 15, 401–48 (1916), and H. Jeffreys, *Cambridge Tracts in Mathematics*, No. 23, Heaviside's operator is treated on the lines of a complex variable and submitted to contour integration. These papers are of very great interest, but it cannot be claimed that they explain the operator.

where t is the variable (always real and positive), p^α the operator, and α and β are any real quantities. In this definition $m!$ satisfies the functional equation

$$f(m) = mf(m-1), \quad f(0) = 1 \qquad \text{......(2)},$$

and in order to fix its value definitely between $0 < m < 1$ Heaviside assumes further that

$$m! = \Gamma(m+1) \qquad \text{......(3)}.$$

Now (2) without any aid from (3) makes $m!$ the product of the first m natural numbers when m is a positive integer, and also infinity when m is a negative integer. Thus whenever α is a positive or negative integer the operator p^α has the same *effect* on a function, as differentiating, or integrating, it α times. This will be true in (1) whatever β may be. This result however does not identify p with the differentiator of the calculus.

Heaviside did not arrive at his definition (1) by taking a formula known to be true for integral values of α, and by assuming it to hold for fractional values. He reached it quite independently from the study of known mathematical solutions of heat diffusion problems*.

We shall use p^α merely to specify precisely a change to be made in a power-term, defined as a function of the type $t^m/m!$, as a result of which the index m is altered in accordance with the equation

$$p^\alpha(t^m/m!) = t^{m-\alpha}/(m-\alpha)! \qquad \text{......(4)}.$$

This definition involves two points in which we must draw a distinction, even when α is integral, between p and the differentiator ($D \equiv d/dt$) of the calculus. These are illustrated by the equations

$$D(at+b) = a, \quad p\left(a\frac{t}{1!} + b\frac{t^0}{0!}\right) = a\frac{t^0}{0!} + b\frac{t^{-1}}{(-1)!}$$

$$D^{-1}a = at + b, \quad p^{-1}\left(a\frac{t^0}{0!}\right) = a\frac{t}{1!} \qquad \text{......(4 a)}.$$

In the first case $t^{-1}/(-1)!$ is numerically zero, and we can disregard it in calculation if all the operations are finished; but if we have still to operate upon it by p^α the quantity really disregarded will be $t^{-1-\alpha}/(-1-\alpha)!$ and this has a calculable value.

In working with p^α we must not add a constant on integration, or differentiate a constant away, since by so doing we add or remove a power term. The operator alters but retains each power term, never introduces a new one, and affects nothing but the index of the power term. If the operand F consists of a group of such terms as

$$F = \sum_s N_s(x^s/s!) \qquad \text{......(5)},$$

the operator simply diminishes by α the index s of the power term, without changing N_s, its numerical coefficient.

* The fullest statement Heaviside gives of his formula, and an example, almost the only one in his book, showing how he arrived at it, will be found in *Electromagnetic Theory*, **2**, 286–90. It should be noted that when he refers to an impulsive quantity as $p.1$ he is using p in a sense quite distinct from that of his formula, and is thinking of his H function to which we shall refer at the close of the present paper.

I shall confine this paper to functions each defined, or expansible, as a group of power terms such as (5); and shall merely use such expressions as e^x, sinh x, etc., as convenient names for such groups. It follows that any equation, or mathematical theorem, denoted by

$$F_1 = F_2 + F_3 + \ldots$$

must ultimately be an equation between two groups of power terms and, since this must hold true for any value of the variable, it must really be an identity. By operating throughout by p^α we merely lower by α the index of each power term, so that the new equation must also be an identity. This will be the case whatever the relative value of the terms so that no divergency question arises. This fact will be completely disguised by grouping the power terms for convenience in the form of functions with special names, and this disguise may in some cases make the action of p^α appear mysterious, but does not affect the precise mathematical character of the operation. Any advantage attending the process is a matter for experience, and not for theory, to determine. An ordinary differential equation, even if its coefficients involve the variable, is merely a case of the superposition of such results, and a differential operator involves nothing more whatever the indices used with the operator p.

This process does not really involve the calculus, and is one of the nature of arithmetic in which we work mechanically to rule without troubling about the theory of the rule. But the process cannot be of use unless two points can be established.

It must be possible, consistently with the assumption about identity of terms, to operate on equations between functions involving infinite series.

It must be possible to establish theorems giving the action of p^α on combinations of functions like (5) so that results can be obtained with sufficient ease to give the process some advantage in the way of convenience over others.

The first point when examined will be found to involve restraints such as those associated with the divergency of series, or with the remainder term in an ordinary expansion. It proves to be a guide in doubtful cases, and not to offer more difficulty than is found with other methods.

The main point is the second, and here we have the whole of Heaviside's work to assure us that the theorems needed can be established. It is true that Heaviside assumed the theorems without proving them, but he subjected his results to such thorough tests that his work is more convincing than formal proofs sometimes are. It is a direct result of definition (4) that p^α when applied to any two functions of type (5) gives us

$$p^\alpha (F_1 + F_2) = p^\alpha F_1 + p^\alpha F_2,$$

and that for any two values α, β

$$p^{\alpha+\beta} F = p^\alpha \,.\, p^\beta F = p^\beta \,.\, p^\alpha F \qquad \text{......(6)},$$

so that p^α and p^β are commutative whatever the values of α and β.

With the help of (4) and (6) some of Heaviside's results can be established without the need of any additional theorem. For consider the differential equation

$$\sigma^2 V = q^2 V \qquad \text{......(7)},$$

where σ is d/dx and q a constant. This equation, though obtained by the calculus, can be looked upon simply as a law of indices, and as a matter of fact this is how such equations are often used. The solution of $(\sigma - q)(\sigma - q) V = 0$ is given as $AP + BQ$ where A and B are constants and P and Q satisfy

$$(\sigma - q) P = 0, \quad (\sigma + q) Q = 0 \qquad \text{......(7 a)}.$$

P and Q are then assumed to be power series in x and the equations are used simply as index relations to determine connections between the coefficients of the series, solutions for which are thus proved to be

$$V = \exp(\pm qx).$$

Now if instead of q^2 being constant we have the case

$$\sigma^2 V = m^2 \frac{dV}{dt} \qquad \text{......(7 b)},$$

a case which covers many important physical problems, we can use exactly the same argument putting $q = mp^{\frac{1}{2}}$. The only difference is that the typical term of the assumed power series instead of being $q^n x^n/n!$ becomes

$$\frac{m^n x^n}{n!} p^{n/2} = \frac{m^n x^n}{n!} \frac{t^{-n/2}}{(-n/2)!}.$$

We use the facts that σ and q are independent; that

$$\sigma . \sigma P = \sigma^2 . P,$$

and that

$$p^{\frac{1}{2}} . p^{\frac{1}{2}} P = p . P = dP/dt.$$

We do not in either case use the calculus except in the form of rule (4) and in the latter case we do not assume that $(d/dt)^{\frac{1}{2}}$ has any meaning, though we do assume, in accordance with (4), that $p^{\frac{1}{2}} P$ has a definite meaning. The solution

$$V = A \exp(qx) + B \exp(-qx), \text{ where } q = mp^{\frac{1}{2}},$$

as given by Heaviside, has no connection with the exponential theorem; only the two exponential series are involved, and these are as definite functions of t as of x, when used, as is assumed, as operators applied to the operand unity, or $t^0/0!$. In the above case we can get another solution by assuming for P and Q

$$P = \sum_{-\infty}^{\infty} \left[T_n \frac{t^n}{n!} + T_n' \frac{t^{n+\frac{1}{2}}}{(n+\frac{1}{2})!} \right] \qquad \text{......(7 c)},$$

with a similar expression for Q, where T_n and also T_n', are ascending power series in x with adjustable constants, which can be found, with the aid of the index equations, so as to satisfy all necessary conditions.

We are not concerned with the manipulation involved, or with the choice of a solution to meet the conditions of a particular physical problem. The point is that

Heaviside's method, as put above, seems in this case to make no assumption which was not originally made by Boole in connection with his symbolical methods. When however we take a more complex case* where

$$q^2 = [r + l(d/dt)]\,[k + s(d/dt)],$$

and where r, l, k, s are constants, we find that Heaviside assumed that he could express q, and any power of q, in terms of his operator p, by substituting p for the differentiator in the above equation and using analysis which is applicable only if p can be treated as an ordinary variable. The power series in p so obtained, in this and similar cases, he interpreted by means of (4); he applied the results to physical problems and obtained solutions which in all cases were confirmed when they could be compared with those got by standard methods. Heaviside was however content with overall tests of final results, and did not test intermediate stages. A proof must be such as to make clear and to justify each step. His results have not been challenged, but Heaviside himself never maintained that these results were strictly proved, however confident he was, and had good cause to be, about their correctness. The convenience of his methods has never been disputed. What is needed to complete the proof is the establishment of theorems justifying the steps of his argument. This leads to the consideration of the action of his operator on products, ratios, roots, etc., of functions like (5), and of similar expressions involving the operator itself. In what follows I shall use the Leibnitz theorem to suggest a fractional operator; show that to make the operator work consistently the Heaviside formula must be used; and use the combination, with the aid of the summation formula for the hypergeometric series, to show that consistent results will be got by applying the operator to the two sides of an equation representing a mathematical theorem.

§ 3. THE INDEX OPERATOR AND LEIBNITZ'S THEOREM

Leibnitz's theorem for the successive differentation of the product of any two functions F and G of x can be regarded as giving the result of the action of an operator determined by F working on the second function G. In its original form this operator can be written

$$p^\alpha F = \sum_{n=0}^{\infty} (1/n!)\, F_n\, [\alpha!/(\alpha - n)!]\, p^{\alpha-n} \qquad \text{......(8)}$$

and in its extended form† as

$$\Pi F = \sum_{n=0}^{\infty} 1/n!\, F_n \Pi_n \qquad \text{......(8 a),}$$

where

$$p \equiv d/dx, \quad F_n \equiv d^n F/dx^n, \quad \Pi_n \equiv d^n \Pi/dp^n,$$

n and α being each positive integers, and Π a function of p with powers restricted to integral values.

Let us adopt the right hand element of (8) as the definition of an operator in terms of a standard operator p^α, where α may have any value. For convenience we shall call it "the *hg* operator" in F.

* See (22) below. † Forsyth, *Differential Equations*, p. 46, 2nd ed. or p. 60, 5th ed.

We propose to consider whether we can so define p^α that p^α and p^β are commutative for all real values of α and β, and are consistent with ordinary differentiations (integrations) when these values are integral ......(9)

and $$p^\alpha F.G = p^\alpha G.F = p^\alpha.FG \quad ......(10).$$

Now without defining p^α let us assume (9) and (10) and take

$$F = x, \quad G = x^m.$$

We shall need to use the relations

$$m! = m(m-1)!, \quad 0! = 1, \quad m! = \Gamma(m+1) \quad ......(11),$$

then if $$Z \equiv p^\alpha(x^m/m!) \quad ......(12),$$

we have by (9) for any integer n

$$p^{\alpha+n}[x^{m+n}/(m+n)!] = p^\alpha.p^n[x^{m+n}/(m+n)!] = p^\alpha(x^m/m!) = Z \quad ...(13).$$

Hence $$Z = p^{\alpha+1}[x^{m+1}/(m+1)!] = [1/(m+1)!]\,p^{\alpha+1}x.(x^m/m!),$$

or by (8)

$$(m+1)Z = [xp^{\alpha+1} + (\alpha+1)p^\alpha](x^m/m!) = (xp+\alpha+1)Z,$$

thus $$(m-\alpha)Z = x(dZ/dx),$$

and we have two values for Z giving

$$p^\alpha(x^m/m!) = x^{m-\alpha}/f(m,\alpha) \quad ......(14),$$

where $f(m,\alpha)$ involves m and α but not x. Now if α is integral we must have whatever m,

$$f(m,\alpha) = (m-\alpha)! \quad ......(15),$$

and this value for $f(m,\alpha)$ will satisfy a number of functional equations readily obtained from (13), or by applying $p^{-\alpha}$ to (14), or by putting

$$\alpha = m \pm n,$$

where n is integral, and operating on $x^m/m!$ by p^α or $p^{\pm n}.p^m$.

Thus Heaviside's definition

$$p^\alpha(x^m/m!) = x^{m-\alpha}/(m-\alpha)! \quad ......(16),$$

follows from (14) and (15), satisfies (9) in all cases, and satisfies (10) for the case taken

$$F = x, \quad G = x^m,$$

because from definition

$$p^\alpha FG = p^\alpha x^{m+1} = (m+1)!\,[x^{m+1-\alpha}/(m+1-\alpha)!],$$

and from (8)

$$p^\alpha F.G = p^\alpha x.x^m = (xp^\alpha + \alpha p^{\alpha-1})\,x^m,$$

which reduces to the same result, while

$$p^\alpha G.F = p^\alpha x^m.x,$$

like other cases shown below, yields a hypergeometric series which can readily be summed by (17) with the same result as above.

Now assuming definition (16) let us take

$$F = x^\beta, \quad G = x^\gamma,$$

where β and γ may have any values, and use (8) to test (10). We have

$$p^{\alpha}F = p^{\alpha}x^{\beta} = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\beta!}{(\beta-n)!} x^{\beta-n} \frac{\alpha!}{(\alpha-n)!} p^{\alpha-n}$$

and $$p^{\alpha-n}.G = p^{\alpha-n}x^{\gamma} = \gamma!\,[x^{\gamma+n-\alpha}/(\gamma+n-\alpha)!].$$

We also have from definition

$$p^{\alpha}.FG = p^{\alpha}x^{\beta+\gamma} = (\beta+\gamma)!\,[x^{\beta+\gamma-\alpha}/(\beta+\gamma-\alpha)!].$$

By substituting in (10) and equating coefficients of $x^{\beta+\gamma-\alpha}$ we have

$$\frac{(\beta+\gamma)!}{(\beta+\gamma-\alpha)!} = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\beta!}{(\beta-n)!} \frac{\alpha!}{(\alpha-n)!} \frac{\gamma!}{(\gamma+n-\alpha)!} \qquad \text{......(17).}$$

If we put this in the form

$$\frac{(\beta+\gamma)!\,(\gamma-\alpha)!}{(\beta+\gamma-\alpha)!\,\gamma!} = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\beta!}{(\beta-n)!} \frac{\alpha!}{(\alpha-n)!} \frac{(\gamma-\alpha)!}{(\gamma+n-\alpha)!} \qquad \text{......(18),}$$

we have on the right hand side a hypergeometric series, and on the left, in accordance with the known summation formula, the sum of this series. These facts may be more apparent if we put

$$a+\alpha = -1 = b+\beta,$$

and $$c = \gamma - \alpha.$$

We then have, after using (11),

$$\alpha!/(\alpha-n)! = (-1)^n\,[(a+n)!/a!]$$

and a similar expression for $\beta!$, etc., so that (18) becomes

$$\frac{(c-a-b-2)!\,c!}{(c-b-1)!\,(c-a-1)!} = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{(a+n)!\,(b+n)!\,c!}{a!\,b!\,(c+n)!}.$$

The ordinary condition for convergence

$$(a+1)+(b+1) < (c+1)$$

becomes in the present notation

$$\beta+\gamma+1 > 0 \qquad \text{......(18 } a\text{),}$$

This condition is independent of α, and only involves the sum of β and γ, which is the index of the power term operated upon. If the series is to be convergent this index must not be -1 or less.

Thus, provided that we use Heaviside's formula in the *hg* operator (8) suggested by Leibnitz's theorem, we can use this operator with consistent results in (10) for the product of any two functions of type x^{β} and x^{γ}. It follows from simple superposition that this must also be true for any two functions of the general type

$$\sum_{s} N_s x^s/s!$$

where N_s is a constant arbitrary for each value of s. Moreover if

$$F = G_1.G_2\ldots,$$

where each factor G is a function of the general type just mentioned, we must get consistent results in working out $p^{\alpha}F$ however the factors are operated upon. We can

expand in many ways. We may multiply all the factors and operate on them term by term; or form the *hg* operator $p^a G$, by taking any factor first, and with the power series in p so obtained operate on the rest term by term; or use the *hg* operator again and so on. But any two series so formed, however complicated they may be, can be shown to have the same sum. No difficulty has been found in proving equality in several most complex cases of this kind.

§ 4. THE INDEX OPERATOR IN USE WITH BINOMIAL AND EXPONENTIAL EXPANSIONS

The harmonious working of p^a, the *hg* operator, and the binomial theorem, can be shown as follows: let

$$F \equiv \sum_{m=0}^{\infty} \frac{\sigma!}{(\sigma - m)!} \frac{a^m x^m}{m!} = (1 + ax)^\sigma \qquad \text{......(19).}$$

If we assume the theorem giving the second expression for F we are entitled to put

$$F_n = \frac{\sigma!}{(\sigma - n)!} a^n (1 + ax)^{\sigma - n} = \frac{\sigma!\, a^n}{(\sigma - n)!} \sum_{s=0}^{\infty} \frac{(\sigma - n)!}{(\sigma - n - s)!} \frac{a^s x^s}{s!}.$$

Now if we put this in the *hg* operator in F we get

$$p^a F . \frac{x^0}{0!} = \sum_{n=0}^{\infty} \sum_{s=0}^{\infty} \frac{1}{n!} \frac{\sigma!}{(\sigma - n - s)!} \frac{a^{n+s}}{s!} \frac{\alpha!}{(\alpha - n)!} \frac{x^{n+s-\alpha}}{(n - \alpha)!},$$

Now put $n + s = m$ with limits 0 to ∞ for m, we get

$$p^a F . 1 = \sum_{m=0}^{\infty} k_m \frac{\sigma!}{(\sigma - m)!} a^m \frac{x^{m-\alpha}}{(m - \alpha)!},$$

where
$$\frac{k_m}{(m - \alpha)!} \doteq \frac{1}{m!\,0!} \sum_{n=0}^{\infty} \frac{1}{n!} \frac{m!}{(m - n)!} \frac{\alpha!}{(\alpha - n)!} \frac{0!}{(0 + n - \alpha)!},$$

the *hg* formula (17) gives $(m + 0)!\ (m + 0 - \alpha)!$ as the sum of this series, so that we find $k_m = 1$ and we get the same result as by using p^a on each term of the F series separately.

Moreover if we expand the *hg* operator in F where F is the series e^{ax} so that F_n is $a^n e^{ax}$ we get on summing by the binomial theorem the power series* in p

$$p^a e^{ax} . G = e^{ax} (p + a)^a G \qquad \text{......(20),}$$

where G is any function of x; and from this if $\Pi(p)$ is any group of power terms in p it follows from superposition that

$$\Pi(p)\, e^{ax} . G = e^{ax} \Pi(p + a)\, G \qquad \text{......(21),}$$

a result also readily following from (8 a) whatever function Π may be of p, provided we can assume that both the binomial theorem and Taylor's theorem apply, and even without the latter assumption if Π is merely a group of power terms in p; indeed the applicability of Taylor's theorem involves the last point. Heaviside made

* The proof indicated assumes that $(p + a)^a$ is expanded in fractional powers of p when a is fractional. We cannot assume without further investigation that we can use as equivalent the expansion of $(a + p)^a$ in integral powers of p.

great use of these theorems in his fractional calculus though he admits that they are only "justified by results." When the powers of p are integral (20) and (21) are of course well established.

§ 5. FUNCTIONS OF THE OPERATOR

It will be found, on referring to the proof of (19), that, although the convergency of the binomial expansion is assumed, subject to this condition, there is no restriction imposed on the value of σ, which may be negative or fractional. The convergency of the series giving k_m is assured by the fact that m is always positive (see (18 a)). It therefore follows from the above that p^α can be used with consistency on all products, ratios, roots and powers of functions of the variable if reducible to type (5), with or without the use of the binomial theorem; always provided that the corresponding expansions are convergent. We have not dealt with cases of corresponding functions of the differentiator, but the above results encourage the belief, due to Heaviside's work, that theorems can also be established for such cases; a result which will complete the justification of his methods.

This justification is simple enough for cases covered by

$$\sigma^2 F = \Pi F \qquad \text{......(22)},$$

where σ is d/dx and Π is any quadratic function of p. Heaviside reduces Π to the form $(p+\rho)^2+c$, and gets rid of ρ by putting

$$F = G \exp(-\rho t),$$

which reduces the differential equation to the form

$$(\sigma^2 - \Delta^2)\, G = 0,$$

where $\Delta^2 = p^2 + c$. This can be dealt with as in a previous case (7 b) provided we can find, and can justify, an operator Δ, which, if applied twice in succession to G, produces the same result as a single application to it of (p^2+c). The natural expression for Δ is $(p^2+c)^{\frac{1}{2}}$, or $p\,(1+c/p^2)^{\frac{1}{2}}$, but even the meaning of such an expression has to be defined when p is Heaviside's operator. Now if the binomial expansion of $(1+y)^{\frac{1}{2}}$ be denoted by $B(y)$ we can define Δ as $pB(y)$ where y is c/p^2. Δ then becomes merely a group of power terms in p with integral indices, and a perfectly definite meaning attaches to such expressions as ΔG, $\Delta^2 G$, etc. All that is needed to justify Heaviside's method is a proof that $\Delta^2 \equiv p^2 + c$. Now this results from assuming the binomial theorem for $pB(y)$, and for its square; but a more convincing form of proof is obtained by multiplying together the two equal series and collecting terms involving y^s, or $(c/p^2)^s$. We then find

$$\Delta^2 = p^2 \sum_{s=0}^{\infty} K_s \left(\frac{c}{p^2}\right)^s,$$

where $s!\,K_s$ is a convergent hg series taking the form given by (17) where $\alpha = s$, $\beta = \gamma = \frac{1}{2}$. This formula thus yields

$$s!\,K_s = 1/(1-s)!$$

and, since s is a positive integer, K_s vanishes for all values of s except 0 and 1. We get

$$K_0 = K_1 = 1$$

and

$$\Delta^2 = p^2\left(1 + \frac{c}{p^2}\right) = p^2 + c.$$

It is curious that in the general case where Π is a quadratic function of p the result is reached without using fractional powers of p; while in the less general case

$$\Delta^2 = m^2 p$$

fractional powers of p are involved.

The only inconsistencies, and these but apparent ones, met with in working with p^α, have been traceable to violation of the formula (4) due to neglect of (4 a) in consequence of the habit of thinking of p as identical in meaning with the differentiator of the calculus. We must not neglect the existence* of a power term like $x^{-m}/(-m)!$ where m is a positive integer, if it has to be operated upon afterwards by p^α, where α is either a fraction or some negative integer equal to or greater than m. As an instance, let F be e^x, sinh x, or cosh x. On using the hg operator we get

$$p^\alpha e^x = (A + B)\, e^x \qquad \text{......(23 } a\text{)},$$

$$p^\alpha \sinh x = A \sinh x + B \cosh x \qquad \text{......(23 } b\text{)},$$

$$p^\alpha \cosh x = A \cosh x + B \sinh x \qquad \text{......(23 } c\text{)},$$

where A and B are each given by

$$\sum_n \frac{1}{n!}\frac{\alpha!}{(\alpha - n)!}\frac{x^{n-\alpha}}{(n-\alpha)!} = \frac{\alpha!\sin\alpha\pi}{\pi}\sum_n \frac{(-1)^n}{n!}\frac{x^{n-\alpha}}{\alpha - n},$$

the values of n being positive and including all even integers for A and all odd integers for B. The second expression is equivalent to the first by a well-known formula. It will be noticed that if α is integral one of the series reduces to zero and the other to unity, as required.

Now if we differentiate equations (23) in the ordinary way we at first expect that (23 a) will reproduce itself and that (23 b) and (23 c) will interchange. This is not actually the case, except that (23 b) becomes (23 c), since (using the second expression for A or B) we get zero as the value of

$$\sinh x \frac{dA}{dx} + \cosh x \frac{dB}{dx}.$$

This expression with A and B interchanged is not zero but a series whose sum can be shown to be

$$-\frac{\alpha!\sin\alpha\pi}{\pi\, x^{\alpha+1}} = \frac{x^{-\alpha-1}}{(-\alpha-1)!},$$

and it follows that if we differentiate the right-hand element of (23 a) or (23 c) we get the expected result as increased by this term.

* When statements are made in print that certain functions of p (only) are not commutative with each other, the conclusion stated is probably traceable to this cause. For two instances see the top of p. 17 and bottom of p. 18 in Cambridge Mathematical Tract No. 23, *Operational Methods in Mathematical Physics*, by H. Jeffreys.

Now
$$p^\alpha p e^x = p^\alpha\left(e^x + \frac{x^{-1}}{(-1)!}\right),$$
$$p^\alpha p \sinh x = p^\alpha \cosh x,$$
$$p^\alpha p \cosh x = p^\alpha\left(\sinh x + \frac{x^{-1}}{(-1)!}\right).$$

So that if we do not dismiss the negative power term we get results quite consistent whichever equation (23) we differentiate.

§ 6. HEAVISIDE'S OPERATORS

Heaviside's operator as defined in (4) does not seem open to criticism. The striking consistency with which it works in conjunction with mathematical theorems is attributable to special properties of the power term (or index function) due to its association with the gamma function. The most obvious characteristic of the power term is that it appears to retain an existence and individuality of its own whatever operations may be performed upon it. The definition of the index operator p^α is very simple except that it involves the gamma function whenever α is fractional (otherwise this function cancels from the equation of definition).

Heaviside's lack of attention to the theory of his operator is striking but characteristic. "It was not the main subject" and he was rarely willing to spend time on side issues. The appeal of his mathematics is due to the fact that he always subordinated it to the problem, arrived at the result required by the shortest route he could find, and used physical conceptions to guide his analysis and prevent it spreading over all the possibilities of a general case. His operator was used as a form of arithmetic with his methods, and these were mostly of his own devising. He would use any known method if convenient, but hardly ever made use of known mathematical results. His method in essence was a mode of getting the formula actually wanted, direct from the differential equation or operator, which he said "not only expresses a problem but also its solution." This should be obtainable from the operator "without ambiguity and without external assistance." He would find a new operator for every fresh type of problem, but would rarely discuss* it after it had served his purpose.

His methods though 40 years old are not widely known or much used except with periodic forces for which they become much the same as more orthodox ones. They are of most value in cases of incomplete diffusion, and, in particular, in the study of the initial waves due to an impulsive disturbance. We select from his examples in each of these cases one to illustrate the method, and to show that it can be justified without the use of unproved assumptions. Incidentally we shall illustrate the use of his impulse function.

* One of the most remarkable operators he ever suggested, and one whose great generality appealed to him very strongly, was (*Electromagnetic Theory*, 2, 348) thrown out merely as a side issue. "Incidentally we get this interesting application of physical to pure mathematics.... The theorem is very striking. It is obviously true by the method employed, being based upon (103) a simple property; but by pure rigorous mathematics there seems nothing whatever about (it) even suggestive of Bessel functions." He plays with this theorem from the middle of one page to the top of the next, but he does not refer to it again.

§ 7. EXAMPLES

We take first the case of a simplified form of Kelvin's earth cooling problem. If the curvature of the earth's surface be neglected, and θ be the temperature at a depth x beneath the surface, the differential equation covering the problem is

$$d^2\theta/dx^2 = (c/k)\, d\theta/dt = q^2\theta,$$

where
$$q = \pm\,(cp/k)^{\frac{1}{2}},$$

t is the time counted from an initial state at which θ has the same value θ_0 for all values of x, and c, k, and g are physical constants assumed known, g being the experimentally measured gradient of temperature at the surface. This is a case covered by (7) and the solution is

$$\theta = \theta_0 e^{-qx},$$

$$-\frac{d\theta}{dx} = q\theta_0 e^{-qx}.$$

At the surface $x = 0$ and we have

$$g = q\theta_0 = \left(\frac{c}{k}\right)^{\frac{1}{2}} p^{\frac{1}{2}}\theta_0 = \left(\frac{c}{k}\right)^{\frac{1}{2}} \frac{t^{-\frac{1}{2}}}{(-\frac{1}{2})!}\theta_0.$$

Now $(-\frac{1}{2})! = \pi$ and Heaviside* takes

$$k = 0{\cdot}0118\, c, \quad 2743\, g = 1, \quad \theta_0 = 4000^\circ \text{ C.},$$

whence t works out to be 103 million years.

We next take the case of a simple one-dimensional problem in which electromagnetic waves travel through the dielectric between the concentric conductors of a cable. The differential equations are

$$-\frac{dV}{dx} = rC + lpC\text{; and } -\frac{dC}{dx} = kV + spV \qquad \text{......(24)},$$

V and C being the voltage and current at distance x, and time t, while r, l, k, s are respectively the resistance, inductance, leakage, and capacity, all per unit length of the cable.

If we define q, ρ, σ, and a velocity v by the equations

$$\left.\begin{array}{ll} slv^2 = 1, & q^2v^2 = (p+\rho)^2 - \sigma^2 \\ 2\rho = \frac{r}{l} + \frac{k}{s}, & 2\sigma = \frac{r}{l} - \frac{k}{s} \end{array}\right\} \qquad \text{......(25)},$$

we get from (24)

$$\frac{d^2V}{dx^2} = q^2V, \quad \frac{d^2C}{dx^2} = q^2C \qquad \text{......(26)},$$

and the solutions as justified by (22) above are

$$V = e^{-qx}V_0, \quad C = e^{-qx}C_0 \qquad \text{......(27)},$$

where $V_0 = F(t)$ is a time function denoting the applied voltage at $x = 0$. C involves both x and t but C_0 involves t only. We suppose the cable to extend indefinitely in

* *Electromagnetic Theory*, **2**, 14.

either direction so that x may be either + or −, but since neither V nor C can increase indefinitely with x the sign chosen for q in taking the root must be such that qx is always +.

Taking first the "distortionless" case with $\sigma = 0$ or $r/l = k/s$ we get for positive x

$$qv = p + \rho$$

$$V = \exp\left[-\frac{x}{v}(p+\rho)\right] F(t) = e^{-\rho(x/v)} e^{-p(x/v)} F(t)$$

$$V = e^{-\rho(x/v)} F\left(t - \frac{x}{v}\right) \qquad \ldots\ldots(28),$$

and for negative x we get the same expression with the sign of x reversed.

Thus we get a positive and a negative wave each moving with velocity v, without distortion, but attenuating as x increases in accordance with the factor $\exp[-(x/v)\rho]$.

For an ordinary cable k will be zero but σ will not vanish. The values of ρ and σ will be the same, but since the attenuation of the wave is determined by the former and the distortion by the latter, we shall retain both symbols in the formulae.

We shall assume that V_0, which $\equiv F(t)$, is constant after $t = 0$ and zero before.

Since $k = 0$ we have from (24) and (27)

$$spV = -\frac{dC}{dx} = -\frac{d}{dx}(e^{-qx}C_0) = qC \qquad \ldots\ldots(29),$$

which at $x = 0$ yields

$$C_0 = s(p/q)V_0 = svV_0(p/qv) \,.\, 1.$$

Now using (20) and (21) Heaviside proceeds

$$\frac{p}{qv}\,.\,1 = \frac{p}{\sqrt{(p+\rho)^2 - \sigma^2}} e^{-\rho t} e^{+\rho t}\,.\,1 = e^{-\rho t}\frac{p-\rho}{\sqrt{p^2-\sigma^2}} e^{\rho t} = e^{-\rho t}\frac{p-\rho}{\sqrt{p^2-\sigma^2}}\frac{p}{p-\rho}\,.\,1$$

$$= e^{-\rho t}\frac{p}{\sqrt{p^2-\sigma^2}}\,.\,1 = e^{-\rho t} I(\sigma t) \qquad \ldots\ldots(30),$$

where $I(\sigma t)$ is defined as the power series in t resulting from the expansion of

$$p/\sqrt{p^2-\sigma^2} = [1 - (\sigma^2/p^2)]^{-\frac{1}{2}}$$

in powers of $(1/p^2)$ transformed afterwards to a function of t by using Heaviside's formula (4). Now this is the Bessel function satisfying the differential equation

$$(\sigma^2 - p^2)I(\sigma t) = (1/t)p\,.\,I(\sigma t) \qquad \ldots\ldots(31).$$

If we change the variable* from t to T where

$$T = t^2$$

* The proof which follows is more like Heaviside than that which he himself gives (p. 307). This for once follows quite conventional lines, and makes much use of known results. It also shows signs of hurry. Few realise the conditions of publication of Heaviside's *Electromagnetic Theory*. These were most extraordinary for an original work of permanent value. The matter was published week after week and year after year in the *Electrician*. No mathematical author was ever granted so much freedom to publish what, when, and how, he liked, and this freedom was granted by a technical newspaper; but not a line could be changed after first publication, since the type was kept set up till the corresponding section of the book was printed. Reference will show that book and

and use P for d/dT, we have

$$\left.\begin{aligned} p &= 2tP \text{ or } 2P = (1/t)p \\ p^2 &= 2P + 2t\frac{dP}{dt} = 2P + 2t^2.2P^2 = 2P + 4TP^2 \end{aligned}\right\} \quad \text{......(32)},$$

so that (31) becomes

$$(\sigma^2 - p^2).I(\sigma t) = 2P.I(\sigma T^{\frac{1}{2}}) \quad \text{......(33)},$$

and this is the value when $n = 1$ of

$$(\sigma^2 - p^2)^n.I(\sigma t) = \{(2n)!/n!\}\, P^n.I(\sigma T^{\frac{1}{2}}) \quad \text{......(34)}.$$

So that to establish (34) we have only to prove it true for $(n + 1)$, assuming it true for n, and to do this we have merely to show that

$$(\sigma^2 - p^2)\, P^n.I(\sigma T^{\frac{1}{2}}) = \{(2n+2)(2n+1)/(n+1)\}\, P^{n+1}.I(\sigma T^{\frac{1}{2}}),$$

and this can readily be done*.

If now we refer to (29) and (30) we have

$$\begin{aligned} C/svV_0 = e^{-qx}C_0/svV_0 = e^{-qx}(p/qv).1 &= e^{-qx}\, e^{-\rho t}\, I(\sigma t) \\ &= e^{-\rho t}\, e^{-(x/v)\sqrt{p^2-\sigma^2}}\, I(\sigma t) \quad \text{......(35)}, \end{aligned}$$

where $I(\sigma t) = (p/\sqrt{p^2 - \sigma^2}).1$.

On expanding $\exp[-(x/v)\sqrt{p^2 - \sigma^2}]$ we get a series containing both odd and even powers of $\sqrt{p^2 - \sigma^2}$. The former will produce zero when operating on $I(\sigma t)$, because, if we substitute for the latter as defined above, we have merely a group of positive integral power terms in p acting on unity as operand. The even powers will yield integral powers of the operator $(\sigma^2 - p^2)$ and with the aid of (34) we get a series of the type

$$\sum_{n=0}^{\infty} C_n P^n.I(\sigma T^{\frac{1}{2}}),$$

which on simplification of C_n will reduce to

$$e^{-(x^2/v^2)P}\, I(\sigma T^{\frac{1}{2}}) = I\{\sigma[T - (x^2/v^2)]^{\frac{1}{2}}\}.$$

So that we obtain the final result

$$C = svV_0\, e^{-\rho t} I\{\sigma[t^2 - (x^2/v^2)]^{\frac{1}{2}}\} \quad \text{......(36)},$$

and this, as in the two cases represented by (28), consists of two travelling disturbances represented in one case by a positive wave and in the other by a negative

journal correspond word for word. Heaviside never lacked matter, but he must sometimes have wished to deal with a fresh point though under great time pressure. New methods may be shorter and better than old ones, but, even in the hands of their author, they are not always so ready for use. The above seems to be a case in point. A device on the same lines as the simple one here adopted is used by Heaviside himself, and with most striking results, only a few pages later (pp. 327–9). He shows how to separate mathematically the positive from the negative wave, and how to follow each independently.

* Since σ^2 is a number $\sigma^2 P^n I = P^n \sigma^2 I$. We can use (33) and (32) to change this expression and also use (32) to change $p^2 P^n I$. Finally if we use Leibnitz's theorem to show that

$$P^n 4TP^2 = (4TP^n + 4nP^{n-1})\, P^2,$$

we obtain the desired relation.

one, each moving with velocity v, and reaching at any time t a limiting distance vt. Of the two time factors, the first represents the attenuation due to ρ, and the second the combined effect of the distortion due to σ and the translation due to v.

§ 8. THE IMPULSE FUNCTION

In the above examples, and in other physical problems in which, according to the usual expression, "a force $F(t)$ is applied at $t = 0$," we can illustrate a use for Heaviside's impulse function. In the physical problem the applied force must always be so stated that it can be calculated for all values of t, although it is supposed to be non-existent for all negative values of t. The analysis gives the solution of a more general problem than the one actually stated. Thus, in (28), V can be calculated not only for negative values of t, but also for values of x numerically greater than vt, though from physical considerations V cannot exist for such values. A similar remark applies to (36). $I(z)$ is a power series in z with only even powers, and thus has a real value even if $z^2 \equiv \sigma^2 [t^2 - (x^2/v^2)]$ is negative. Heaviside maintained that his operational methods were precise in every respect in regard to the particular problem. He regarded the forces as applied to a physical state in which all the fluxes are originally zero, but adjusted to the initial conditions of the problem by the action of these forces during a short starting interval before the actual commencement of the stated problem. His examples in most cases were concerned with moving electromagnetic waves and with the results of their superposition. He was especially interested in what occurred near the wave front.

He made use of an impulse function of the time which we shall denote by $H(t)$. He does not appear to define it precisely but his use of it is consistent with the following statement.

The value H is either zero or unity except in the close neighbourhood of certain times t_1, t_2, etc. at which it changes from one of these values to the other. The change is brought about continuously during a short interval of time $(2c)$ from $t - c$ to $t + c$, and during this interval it may have any unknown or arbitrarily chosen values consistent with being always real, single-valued and continuous, and with having the right values at the limits. Usually there is only one time $t = 0$ at which H changes. $H = 0$ if $t \leqslant -c$; $H = 1$ if $t \geqslant +c$; and H is arbitrary if $-c < t < +c$.

Now in theory a Fourier series can be found to fulfil these conditions, however small c may be compared with the unit in terms of which t is measured. H is ultimately a power series in t with positive integral indices. It is real, single-valued and continuous, and can always be differentiated and integrated.

Each applied force given as $F(t)$ will be associated with a corresponding $H(t)$ function, such that HF represents the applied force for all values of t. The initial conditions can be assumed to result from applying forces such as HF during the starting interval $2c$ to a state in which all the fluxes are originally zero. For the problem itself t is always positive and greater than c. By Laurent's theorem any $F(t)$ can be expanded in a power series in t provided t is similarly restricted. Finally we can assume c indefinitely small, and that F, H, and the product FH are each power series in t.

If now we refer to (27) and (28) and use HF instead of F for the applied force, we get

$$V = e^{-\rho . x/v} F[t - (x/v)] H[t - (x/v)],$$

a quantity which necessarily vanishes not only when t is negative, but also when x exceeds vt. Similarly in case (29) we must use $V_0 H$ instead of V_0 and we only arrive at (30) on the assumption that H is unity, which means that t must be positive; but even so H is always a function of t and as such is present in (30). Equation (34), while true when the operators on the two sides are used as equivalent ones as applied to I, cannot be assumed true when the operand is HI unless H can be assumed to be unity throughout. Each side must vanish if t is negative, but there appears no obvious way of proving that this is so if x exceeds vt. It can however be shown* that the operator applied to HI in (35) can be put in the form

$$e^{-\rho t} e^{-(x/v)(p^2-\sigma^2)^{\frac{1}{2}}} = e^{-\rho t} e^{-(x/v)p} f(1/p),$$

where $f(1/p)$ only involves negative integral powers of p.

The operator is thus compounded of three successive operators, one of which attenuates, the next translates, and the third distorts the disturbance denoted by HI.

The coefficients in the power series $f(1/p)$ involve v, x, and σ but do not contain the variable t. We thus have from (35) if x be positive

$$\begin{aligned} C/svV_0 &= e^{-\rho t} e^{-(x/v)p} f(1/p) H(t) I(\sigma t) \\ &= e^{-\rho t} f(1/p) H[t - (x/v)] I\sigma[t - (x/v)], \end{aligned}$$

which must vanish whenever t is negative or whenever x exceeds vt. If x be negative a similar relation holds with the sign of x changed throughout.

The result is not to invalidate (36) but to show that this equation only holds for values of t and of x for which

$$t > +c \text{ and } |x| < vt.$$

The above illustration of the use of the impulse function was not actually given by Heaviside, though he possibly regarded it as obvious. It was chiefly in problems involving progressive waves, and the superposition of their effects when they are subjected successively to reflections at boundaries that Heaviside made use of his function. He showed clearly the connection between a moving wave solution and one involving stationary waves as represented by a Fourier series. His work in this connection supplies a physical interpretation of the various formulae which is of great value to anyone interested in physics.

DISCUSSION

Dr T. J. I'A. BROMWICH (communicated): The statements made by Dr Sumpner (pp. 404, 405) on the general character of Heaviside's methods of working with operational symbols, must be accepted by all students of the subject: but I venture to question the first footnote (on p. 405), with reference to the use of complex integrals as a means of interpreting Heaviside's operators. Speaking for myself, I

* See *Electromagnetic Theory*, **2**, 316–21 (1899). In fact we have $\log f = (x/v) p \{1 - [1 - (\sigma^2/p^2)]^{\frac{1}{2}}\}$, which shows that $\log f$, and therefore also f itself, consists of a power series in $1/p$.

certainly *do* claim that this interpretation explains the success of Heaviside's work in general: and further that the complex integrals throw light on some points not easily explained in any other way.

Of course no one would suggest that Heaviside himself used complex integrals as an aid to his own discoveries: but his very phrases quoted by Dr Sumpner, seem to indicate that he would have been the last to tie down subsequent workers by insisting on the exclusive use of his own methods. I would add that we were in correspondence on this subject at intervals for some years, starting in 1914; and that he appeared glad to know that a foundation of pure mathematics had been worked out, so as to provide firm footing for the bold and experimental methods of the adventurous pioneer.

Further, the success of Heaviside's methods was due, in a large measure, to his extraordinary power or instinct for detecting hidden sources of error, which might easily lead to unsound conclusions: whereas, if the operators are interpreted in the form of complex integrals, the slippery places reveal themselves to almost any worker with a competent knowledge of modern analysis.

I proceed to illustrate these remarks by some examples; it will be convenient to state at once, without proof, that the complex integral which interprets an operator $f(p)$ is*

$$\int_{c-\iota\infty}^{c+\iota\infty} \frac{d\lambda}{2\pi\iota\lambda} f(\lambda)\, e^{\lambda t},$$

where $c > 0$.

It is then fairly easy to prove that if $f(p)$ is an algebraic function, and is expanded in descending powers of p, the derived series, which will be in ascending powers of t, will be convergent, however large t may be. On the other hand, it is usually found in numerical work, that a definite time-interval τ exists, such that when $t/\tau > 10$, say, the convergence of the series becomes intolerably slow; so that other means of calculation become essential. Heaviside then expands $f(p)$ in ascending powers of p; and the derived series will proceed in powers of τ/t but will not converge. Such series can be proved, from the integral formulae, to have an asymptotic character†, so that the error in stopping at any stage is of the same order as the following term or less.

Ex. 1. (*Electromagnetic Theory*, **2**, § 242.) *Coil at the end of a cable.*
Here $f(p)$ is of the form $1/\{1+\sqrt{(\tau p)}\}$: the series

$$\frac{1}{(\tau p)^{\frac{1}{2}}} - \frac{1}{\tau p} + \frac{1}{(\tau p)^{\frac{3}{2}}} - \frac{1}{(\tau p)^2} + \dots$$

leads to the (always convergent) series

$$2\sqrt{\left(\frac{t}{\pi\tau}\right)}\left\{1 + \frac{2}{3}\frac{t}{\tau} + \frac{2.2}{3\cdot 5}\left(\frac{t}{\tau}\right)^2 + \frac{2.2.2}{3\cdot 5\cdot 7}\left(\frac{t}{\tau}\right)^3 + \dots\right\} - (e^{t/\tau} - 1).$$

* When $f(p)$ contains fractional powers of p, the corresponding terms in λ are made one-valued by supposing a cut along the negative part of the real axis: and such powers are positive (and real) on the positive part of the real axis, in the λ-figure. Also the path of integration passes on the positive side (the right-hand, in the usual diagram) of $\lambda = 0$. See *Proc. Lond. Math. Soc.* (2) **15**, 401, §§ 4, 5 (1916); and *Proc. Camb. Phil. Soc.* **20**, 411, § 4 (1921).

† A proof is given, for one type, at the end of the second paper just quoted: the reader will have no difficulty in carrying out a similar investigation in other problems.

And the alternative form

$$1 - (\tau p)^{\frac{1}{2}} + (\tau p) - (\tau p)^{\frac{3}{2}} + (\tau p)^2 \dots$$

gives the non-convergent asymptotic series

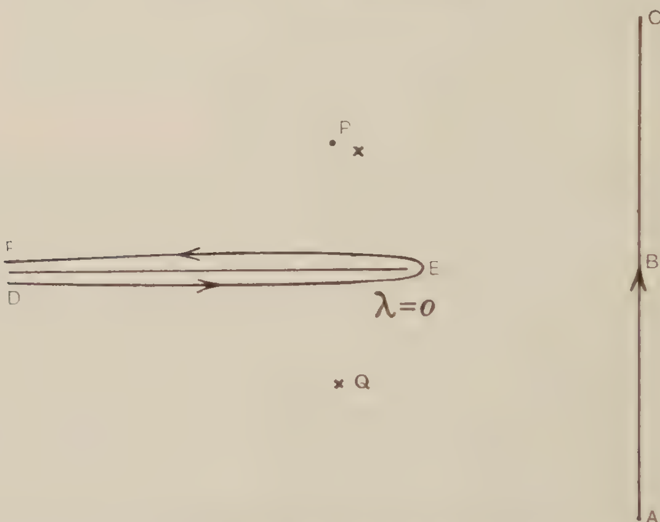
$$1 - \sqrt{\left(\frac{\tau}{\pi t}\right)} \left\{1 - \frac{t}{2\tau} + \frac{1.3}{2.2}\left(\frac{t}{\tau}\right)^2 - \frac{1.3.5}{2.2.2}\left(\frac{t}{\tau}\right)^3 + \dots\right\}$$

from which numerical results are easily found when $t/\tau \geqq 10$.

From the point of view of the complex integral, the first step (expanding in descending powers of $\lambda\tau$) requires that $\lambda\tau$ shall be greater than 1 (in absolute value): this is satisfied by placing the original path ABC sufficiently far to the right (by taking in fact $c\tau > 1$). The values of the integrals are found from results originally given by Cauchy: but undoubtedly the most rapid method of reproducing their values is to use Heaviside's formulae*

$$p^{\frac{1}{2}} = \frac{1}{\sqrt{(\pi t)}}$$

and to operate on this by integration or differentiation.



On the other hand, to obtain the asymptotic series, the path of integration is first bent round from ABC to DEF in the diagram; then we can write $\lambda\tau = -\mu^2$, $\sqrt{(\lambda\tau)} = +\iota\mu$ on the upper path EF, $\sqrt{(\lambda\tau)} = -\iota\mu$ on the lower path DE, and the complex integral becomes

$$1 - \frac{2}{\pi}\int_0^\infty \frac{d\mu}{1 + \mu^2} e^{-\mu^2 t/\tau},$$

where the 1 arises from the integral round $\lambda = 0$.

* An investigation of this formula from the point of view of complex integrals is carried out in § 4 of the second paper quoted. I think there can be no doubt that Heaviside himself regarded this result as the key to his subsequent developments: from his letters, it would seem that the investigation given in *Electromagnetic Theory*, 2, 287–8, is substantially his original method. It will be noted that the integrals used there are the same as those obtained at once from the complex integrals.

Then, if we are content with, say, three terms of the asymptotic series, we may write

$$\frac{1}{1+\mu^2} = 1 - \mu^2 + \mu^4 - \frac{\mu^6}{1+\mu^2}$$

so that the error introduced by using only three terms in the series* is

$$\frac{2}{\pi}\int_0^\infty \frac{\mu^6\,d\mu}{1+\mu^2}\,e^{-\mu^2 t/\tau} < \frac{2}{\pi}\int_0^\infty \mu^6\,d\mu\,e^{-\mu^2 t/\tau}$$

which is now seen to be the fourth term of the asymptotic series itself.

At first sight it may seem strange that in this method of calculation no trace is left of the terms in Heaviside's series which contain integral powers of p (that is, τp, $(\tau p)^2$, etc.): the explanation is that these terms contribute equal amounts to the integrals along DE, EF, and these terms accordingly cancel out in pairs. It is in this sense that we must interpret the formulae

$$p\,.\,1 = 0, \quad p^2\,.\,1 = 0, \quad p^3\,.\,1 = 0, \text{ etc.}$$

which usually occur only in asymptotic series.

Let us consider next another example, in which more care is needed.

Ex. 2. (*Electromagnetic Theory*, **2**, § 244.) $f(p) = 1/\{1 + (\tau p)^{\frac{3}{2}}\}$.

The series in ascending powers of t/τ, is derived from

$$\frac{1}{(\tau p)^{\frac{3}{2}}} - \frac{1}{(\tau p)^3} + \frac{1}{(\tau p)^{\frac{9}{2}}} - \frac{1}{(\tau p)^6} + \dots$$

and is

$$\frac{4\pi}{3}\left(\frac{t}{\pi\tau}\right)^{\frac{3}{2}}\left\{1 + \frac{2.2.2}{5\cdot7\cdot9}\left(\frac{t}{\tau}\right)^3 + \frac{2.2.2.2.2.2}{5\cdot7\cdot9.11.13\cdot15}\left(\frac{t}{\tau}\right)^3 + \dots\right\}$$
$$- \left\{\frac{1}{3!}\left(\frac{t}{\tau}\right)^3 + \frac{1}{6!}\left(\frac{t}{\tau}\right)^6 + \frac{1}{9!}\left(\frac{t}{\tau}\right)^9 + \dots\right\}.$$

Heaviside gives a short table derived from this series from $t/\tau = \frac{1}{2}$ to $t/\tau = 6$.

But if we simply expand $f(p)$ in the form

$$1 - (\tau p)^{\frac{3}{2}} + (\tau p)^3 - (\tau p)^{\frac{9}{2}} + \dots$$

the derived asymptotic series is not sufficient to express the value of the operator. The result given by Heaviside is

$$1 + \frac{\pi}{2}\left(\frac{\tau}{\pi t}\right)^{\frac{3}{2}}\left\{1 - \frac{3\cdot5\cdot7}{2.2.2}\left(\frac{\tau}{t}\right)^3 + \frac{3\cdot5\cdot7\cdot9.11.13}{2.2.2.2.2.2}\left(\frac{\tau}{t}\right)^6 - \dots\right\} - \tfrac{4}{3}e^{-t/2\tau}\cos\left(t\sqrt{3}/2\tau\right)$$

where the presence of the last term could not be naturally foretold without further investigation†.

The explanation of the extra term, from the point of view of the complex integral, arises from the fact that the denominator $1 + (\lambda\tau)^2$ vanishes at two points

* Starting from the known formula

$$\frac{2}{\pi}\int_0^\infty d\mu\,e^{-\mu^2 t/\tau} = \frac{1}{\sqrt{(\pi t/\tau)}}$$

which is equivalent to Heaviside's $p^{\frac{1}{2}} = 1/\sqrt{(\pi t)}$, the reader may apply the process of differentiation to deduce that the following terms agree with Heaviside's calculations from $p^{\frac{3}{2}}$, $p^{\frac{5}{2}}$, etc.

† Heaviside's investigation will be found in the same volume, § 448: his earlier methods are given in *Electrical Papers*, **1**, 153–169.

(P, Q in the diagram) corresponding to $\lambda\tau = e^{2\pi i/3}$ or $e^{-2\pi i/3}$: and thus the integral along ABC is equal to that along DEF, together with the sum of the residues at P, Q: and these, when combined into a real form, provide the correction found by Heaviside*. The actual manipulation of the asymptotic series proceeds on lines similar to those adopted in the former example.

These two examples will probably suffice to show how the complex integral acts as a foundation for the symbolic method: on the other hand, the operational process is the more rapid and compact, serving in fact as a kind of shorthand for these more elaborate calculations. I pass to some other points arising out of Dr Sumpner's paper.

In the solution (§ 7) of the equation

$$\frac{\partial^2\theta}{\partial x^2} = \frac{c}{h}\frac{\partial\theta}{\partial t} = q^2\theta, \quad \text{if } q = \sqrt{(cp/h)},$$

it is necessary to distinguish clearly between the two solutions e^{-qx} and e^{+qx}, the former being the correct form if x can extend to $+\infty$. But, from the operational point of view, it seems out of the question to attach any particular sign to the symbol q; on the other hand, when q is regarded as representing a multiple of $\sqrt{\lambda}$ (in the complex integral), the symbol q will have the same algebraic character as $\sqrt{\lambda}$. Now, in the diagram, $\sqrt{\lambda}$ will be real and positive at the point B; thus $\sqrt{\lambda}$ will have a positive real part at any point along the path ABC (for $\sqrt{\lambda}$ will vary continuously and its real part will vanish only when λ is real and negative). Thus we may treat q as having a positive real part in the formation of our solutions; this will not prevent us from manipulating the path from ABC to DEF, when we require to transform the integrals.

Again, in Heaviside's general cable problem, we are led to the formula for q

$$q^2v^2 = (p + \rho)^2 - \sigma^2,$$

as in equations (25) of Dr Sumpner's paper; and again the distinction between e^{-qx} and e^{+qx} is required. Here, however, the cut required (to give a single value for the square-root) does not extend along the whole negative axis, but simply between the points given by $\lambda = -(\rho + \sigma)$ to $\lambda = -(\rho - \sigma)$. Then the value of the square-root (which is positive at the point B) is easily seen to be approximately

$$\lambda + \rho - \sigma^2/2\lambda$$

when $|\lambda|$ is large. Thus again we arrive at the form e^{-qx}: and when the corresponding complex integral is examined, the exponential term is approximately

$$\exp\left\{\lambda\left(t - \frac{x}{v}\right) - \frac{\rho x}{v} - \frac{\sigma^2 x}{2\lambda v}\right\}.$$

It follows that, when $vt < x$, the path ABC can be bent out to the right instead of the left, and then the value of the integral is zero; this simple argument replaces

* Another illustration is given by Heaviside's generalised cosine series (*Electromagnetic Theory*, 3, 198, 218): further examples are given at the second place, one of which, (35), is not quite correct. This was noted by Heaviside as being then "in an experimental stage." The necessary alteration is easily deduced from the complex integral, but need not be written out here.

that given in Dr Sumpner's § 8. When $vt > x$, the path of integration may be bent round to the left, and may conveniently be transformed into an ellipse enclosing the cut. The evaluation of the result in Heaviside's formula containing

$$I_0 \{\sigma \sqrt{(t^2 - x^2/v^2)}\}$$

is not difficult, but requires some preliminary evaluations of definite integrals which would occupy too much space here*.

Finally, with reference to Heaviside's impulse function of the time (see § 8 of Dr Sumpner's paper) it may not be out of place to remark that the simplest analytical formula for the function is

$$H(t) = \int_{c-\iota\infty}^{c+\iota\infty} \frac{d\lambda}{2\pi\iota\lambda} e^{\lambda t}, \quad (c > 0)$$

which is unity when t is positive, and zero when t is negative†. This formula is, in fact, given by writing $f(p) = 1$ in our general statements and it may be converted into a form of Fourier's integral‡ (not a series).

AUTHOR's reply: I am glad that my paper has been the cause of such an interesting and valuable communication from Prof. Bromwich. As regards the footnote on page 405, I quite admit that the Bromwich "interpretation explains the success of Heaviside's work in general" and that it illuminates the whole subject. It seems to me, however, that the steps of Heaviside's method still need more explanation than is to be found either in the papers of Bromwich and Jeffreys, who use complex integrals, or in that of J. R. Carson who uses integral equations and restricts his analysis to the field of real quantity. These papers, in my view, set forth independent methods of establishing results such as those of Heaviside, rather than a detailed justification of the Heaviside method in itself.

* One method is to show that the result depends only on $(t^2 - x^2/v^2)$, by a simple change of variable in the definite integral. The value at $x = 0$ is found most quickly by Heaviside's own analysis to involve $I_0(\sigma t)$.

† More details on this point will be found in the tract by Dr H. Jeffreys (*Operational Methods in Mathematical Physics*, Cambridge, 1927): see in particular pp. 21–25.

‡ The result (due to Cauchy) is found by bending the path ABC to the right ($t < 0$), or left ($t > 0$): in the latter case the residue at $\lambda = 0$ has to be included in the formula.

A HYGROMETER FOR USE IN TIMBER-SEASONING KILNS

By EZER GRIFFITHS, D.Sc., F.R.S.
Physics Department, The National Physical Laboratory, Teddington, Middlesex

Received February 26, 1929. *Read May* 24, 1929

ABSTRACT. The hygrometer is of the wet and dry bulb type using mercury-in-steel thermometers. The bulbs and capillaries are enclosed in a steel tube which can be inserted through a hole 3 in. in diameter and explorations up to 12 ft. can be made with it. One dial serves both thermometers, the two pointers being arranged so that one moves over the top face of the dial and the other is brought round from beneath on to the outside edge. An electrically operated fan draws air past the two bulbs. The moistening of the covering over the wet bulb is effected in two ways: (*a*) by a reservoir, (*b*) by connexion to an external water supply.

In many industrial processes it is necessary to vary the humidity of the air in a chamber in accordance with a definite schedule.

For timber-drying kilns the common practice has hitherto been to employ the unventilated type of compound spiral thermograph and to carry out the process according to empirical rules worked out as the result of observations with a particular kiln. For the scientific study of timber drying it is, however, essential to obtain data which can be interpreted in terms of definite physical concepts such as relative humidity, temperature, rate of air circulation, etc.

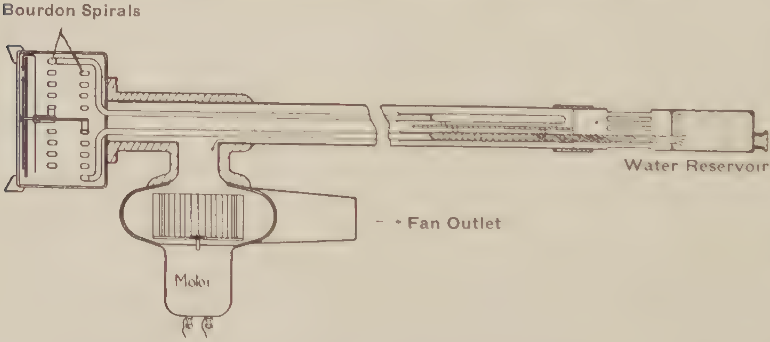


Fig. 1.

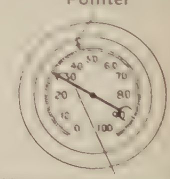


Fig. 2.

The hygrometer shown in Fig. 1 was designed for experimental work in which it was necessary to measure humidities at various distances from the kiln face up to 12 ft. from the walls, which were provided with a series of port-holes. The instru-

ment is of rugged construction, the thermometric elements consisting of two mercury-in-steel thermometers indicating on one dial. The two pointers are arranged as shown in Fig. 2. The steel bulbs and capillary tubes are fixed in a light steel tube 12 ft. in length and perforated at its lower end. Near the point where the tube is attached to the dial an electrically operated fan is fitted on to a branch tube. This fan draws air from the kiln past the thermometer bulbs and capillaries and discharges out to the atmosphere. In the choice of fan consideration must be given to its suitability for this purpose as regards electrical insulation, for moisture is liable to condense in the fan casing and, if it has access to the windings, trouble soon arises.

The wet bulb is sheathed with a layer of muslin which has been treated so as to remove all grease and filling matter. Between the muslin and the bulb a series of fine wires, equally spaced around the bulb, are laid. At the lower end the wires are bunched together and pass into the water reservoir. The function of these wires is to distribute the water supply evenly beneath the covering of the wet bulb. The reservoir was fitted with a prolongation tube leading right up to the bulb; this contained a wick to feed the water along the bundle of wires and the muslin.

The above-described method of keeping the bulb covering wet was found satisfactory for general work. At high temperatures and low humidities it was found advisable to connect the feed pipe to a small-bore tube extending out of the kiln. The tube was connected to a reservoir from which a definite stream of water could be delivered by adjustment of the air pressure until sufficient water had flowed to cause occasional drips to fall from the bulb covering. By the aid of a telescope the condition of the bulb covering could be observed.

The author wishes to thank Mr A. Snow, Observer in the Physics Department, for his assistance.

REVIEWS OF BOOKS

Practical Primary Cells, by A. MORTIMER CODD. Pp. viii + 127, with 53 diagrams. (London: Sir Isaac Pitman and Sons. Ltd.) 5*s*.

The author, with a wide practical experience of his subject, aims to give within moderate compass working details of the primary cells in most general use in industry and in the laboratory. Chief place is fittingly given to the Leclanché cell, to which over a quarter of the book is devoted. We may read of the Lalande, a cell extensively used in railway work, and of cells less familiar, such as the Bleecke-Love and the Velvo-carbon; and at the end a table of information is supplied in regard to 172 forms of primary cell, with electromotive forces ranging from 0·5 up to 3·14 volts.

In addition to treating the generally recognised characteristics of cells the author directs attention to the practical consequences of the less well known phenomena of endosmosis and the Becquerel effect. The superiority of the single-fluid over the double-fluid cell is emphasised, and preference is expressed for the choice of ferric chloride as a depolariser. We learn with astonishment that primary cells are employed for charging secondary cells!

In respect of the theory of the action of the cell the treatment seems open to criticism. The electro-chemical series is given in the opening chapter, and the impression tends to be conveyed that the electromotive force of a cell is a mere matter of the elements constituting the electrodes. But there remains the important part played by the liquids of the cell; it is possible, for example, to construct a cell in which copper is the negative electrode and zinc the positive, so that copper goes into solution and zinc is deposited as the cell supplies current. Exception must also be taken to the appearance of the obsolete Grotthus hypothesis. In making these remarks, however, it is fair to add that the theory of the voltaic cell presents many difficulties, which hitherto have not been satisfactorily resolved.

D. O.

Primary Physical Science, by WILLIAM R. BOWER, B.Sc., A.R.C.S. Pp. 302 with 153 diagrams. (London: Sir Isaac Pitman and Sons, Ltd., 1928.) 5*s*. net.

Many attempts have been and are being made to write science text-books suitable for the lower forms in secondary schools, for central or modern schools, and for children attending evening classes and technical institutes, but a real success has not yet been achieved. One factor which has determined this lack of success is that writers have offered a form of watered-down secondary school book for children in the other types of school, ignoring the fact that a central or modern school education is not inferior to a secondary school education, but is essentially different from it; the two are not to be measured in the same units.

The book at present under review is well printed on good paper and excellently bound, but the subject-matter and method of treatment belong to the period when science teaching in schools was in its infancy, even though an attempt is made by the author to bring the book up to date by the inclusion of a diagram of the crystal structure of common salt, an account of the internal combustion engine, and some historical notes. Hydrogen is introduced by way of zinc and sulphuric acid, and the properties of gases are printed in heavy type to encourage learning by heart. The introduction to physics is given in a chapter on measurement in which the reader is invited to remember a considerable

umber of relations of which "the fluid ounce equals 28·35 cubic centimetres" is typical. n the chapter on heat, temperature is defined in terms of heat in spite of Preston's brilliant eatment in which he shows that temperature, being a percept, is more fundamental an heat, which is a concept.

This is what may be described as a "fact book," and if we judge it by the standards f its type it is good; doubtless it will be successful as a cram book for some elementary xaminations, but we feel that even though a child may be filled with scientific facts yet e may still know little or nothing about science. The inclusion of science in the curricula f elementary and central schools cannot be justified if this is science. In the new scheme f education which is developing throughout the country we plead for a true presentation f science with special attention to its relation to the life of the child and of the community; e need first a selection of suitable material and then a presentation suited to the mental evelopment of the child. The consultative committee appointed by the Board of Education, under the chairmanship of Sir W. H. Hadow, has indicated in its recently published eport that there is a distinct lack of good first-stage science text-books, but the book nder review appears to fall short of the style of text-book which the committee hope will ppear, in due course, to fill the gap which they have disclosed.

V. T. S.

Principles of Radio Communication, by Prof. John H. Morecroft, assisted by A. Pinto and Prof. W. A. Curry. Second edition, thoroughly revised. Pp. xiv + 1001. (New York: John Wiley and Sons, Inc.; London: Chapman and Hall, Ltd., 1927.) 37*s*. 6*d*. net.

The science and practice of radio communication have developed so rapidly during he past ten or fifteen years, and now cover such a wide field, that it has become impossible o confine an adequate technical description of the subject within the covers of a single olume. Whatever alteration may take place on the practical engineering side, however, t is unlikely that the fundamental principles of the generation, transmission and reception of electromagnetic waves will be materially changed; and it is with such principles that the somewhat formidable volume compiled by Prof. Morecroft is concerned. It is the second edition of a book with which many radio engineers are already familiar; and although, as mentioned in the preface, two chapters from the first edition have been deleted and a third has been considerably reduced in length, the amount of new material dealt with is such as to bring the present volume up to 1001 pages.

The first chapter of the book describes in a very clear manner the fundamental ideas of the flow of electrons in conductors, the production of electric and magnetic fields and the phenomena associated with the passage of alternating currents in resonant circuits. The electrical properties of a typical radio-frequency circuit, viz. resistance, inductance and capacity, are ably dealt with in the next chapter which concludes with a sound but brief discussion of the problem of screening one circuit from another, for both electric and magnetic fields. Here, as in other portions of the book, there is a very noticeable absence of mention of important British research work, such as that of Butterworth and Howe, on the resistance of coils and the capacity of aerials respectively. Chapter III entitled "Laws of Oscillating Circuits" deals with a branch of the subject which is fundamental to either transmission or reception in radio; but a large portion of the chapter is concerned with damped oscillations, which are rapidly becoming of academic interest only. Further, the fact that Chapter V, entitled "Spark Telegraphy," also deals with an obsolescent portion of the subject suggests that a more drastic curtailment of these chapters might have been carried out. The space so saved might with advantage have been devoted to the intervening fourth chapter which is all too short in its attempt to review the phenomena associated with atmospherics and the propagation of waves over the earth's surface in only forty pages. The sixth chapter is the largest in the book and

may be regarded as a small monograph on the application of thermionics to modern radio communication. In its description of the use of the valve as a rectifier, amplifier and oscillator the chapter is very sound and fairly up-to-date, but it might have been sub-divided and re-arranged with some advantage. In this connection it is noticeable that some repetition occurs in the following chapter, which is entitled "Continuous Wave Telegraphy." A good description of the principles of modulating continuous oscillations for use in radio telephony and broadcasting is contained in Chapter VIII. The concluding chapters entitled "Antennae and Radiation" and "Amplifiers" could possibly have been better placed earlier in the book; but otherwise deal usefully with the subjects concerned.

The book is conspicuous throughout for the large amount of experimental data supplied and the reproduction of many excellent oscillograph records illustrating the various phenomena under discussion. These obviously represent the results of the author's own research work, and they form a valuable addition to a text-book which must form a most useful reference volume for every serious technical worker in the field of radio communication. Finally, if the author ever contemplates the somewhat arduous task of compiling a third edition it may be suggested first that a saving of space could be effected by a re-arrangement which avoids certain overlapping between chapters, and also that the value of the book will be increased if more attention is paid to the work carried out in Europe during recent years.

R. L. S. R.

The National Physical Laboratory: Report for the year 1928. Pp. vi + 284. (H.M. Stationery Office for the Dept. of Scientific and Industrial Research.) 9*s*.

This report summarises a year of solid achievement and one in which some important developments have taken place. The first of these concerns the new physics building for which plans were drawn up in 1924 only to be shelved for the time being. We are glad to learn that the contract for the central block of the building has now been placed, and that construction has already begun. This first instalment, when completed, will be occupied by the heat division for which improved conditions are much required. The second is the completion of the high tension building and the erection of the million volt transformers. The report contains a brief description of the building and its plant, and a most striking reproduction of a photograph of the million volt spark. A third is foreshadowed, namely the erection of a compressed air tunnel to meet requirements in aeronautics. This will be begun within the next few months, and its use will enable results obtained from aircraft models to be applied directly to the full-scale machine without the necessity of a scale correction.

In a brief review it would be invidious to single out any one division for special notice. Each year brings its increase in the scope of work upon which the Laboratory is engaged and in the extent to which manufacturers are seeking its assistance. Special reference, however, may be made to the very full programme of work relating to fundamental standards which has been completed and to the increasing volume of experimental work in view under this head. This is work in which the Laboratory has always been a pioneer and which indeed could not be carried out anywhere else in this country.

Lastly, one may hope with the Chairman that a suitable building may soon be available for the sound division, the work of which is susceptible to extraneous disturbances unless specially constructed rooms are available. To those of us who dare not count up the hours we have spent in endeavours to hear lectures and speeches in faulty buildings, the subject of architectural acoustics is one that is very close to our hearts.